Essays in High Frequency Trading and Market Structure

Michael Harrison

A thesis submitted in partial fulfilment of the requirements of the University of East London for the degree of Doctor of Philosophy

March 2022

Abstract

High Frequency Trading (HFT) is the use of algorithmic trading technology to gain a speed advantage when operating in financial markets. The increasing gap between the fastest and the slowest players in financial markets raises questions around the efficiency of markets, the strategies players must use to trade effectively and the overall fairness of markets which regulators must maintain. This research explores markets affected by HFT activity from three perspectives. Firstly an updated microstructure model is proposed to allow for empirical exploration of current levels of noise in financial markets, this illustrates current noise levels are not disruptive to dominant trading strategies. Second, a ARCH type model is used to de-compose market data into a series of traders working price levels to demonstrate that in cases of suspected market abuse, regulators can assess the impact individual traders make on price even in fast markets. Finally, a review of various HFT control measures are examined in terms of effectiveness and in light of an ordoliberal benchmark of fairness. The work illustrates the extents to which HFT activity is not yet disruptive, but also shows where HFT can be a conduit for market abuse and provides a series of recommendations around use of circuit breakers, algorithmic governance standards and additional considerations where assets are dual listed in different countries.

Table of Contents

Definition

In this thesis, the term ‘high-frequency trading’ is used extensively. This term is commonly abbreviated as HFT. The present study uses the abbreviation HF to refer to ‘high frequency’ and HFT to refer to high-frequency trading.

An HFT, in general terms, is an automated trading system that uses lines of code that are designed to operate as quickly as possible. It is not possible to state that a certain system speed is HF; rather, it is more likely that certain characteristics are likely to define a system as an HFT, such as the different strategies, the order types, the use of latency-minimising technology, and the application of an inventory-neutral strategy. It is also noteworthy that the latency element of HFT has varied over time, as this is an evolving area of activity, and there is a strong incentive to speed up and overtake others in the market.

List of abbreviations

ARCH	Autoregressive Conditional Heteroscedasticity
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
MiFID	Markets in Financial Instruments Directive
NYSE	New York Stock Exchange
NASDAQ	National Association of Securities Dealers Automated Quotations
AT / ATS	Algorithmic Trader / Algorithmic Trading System
LSE	London Stock Exchange
HKEX	Hong Kong Stock Exchange
OTR	Order to Trade Ratio
ESMA	European Securities and Markets Authority
SEC	Securities and Exchange Commission
CFTC	Commodity Futures Trading Commission
FCA	Financial Conduct Authority
PRA	Prudential Regulatory Authority
FSA	Financial Services Authority
IMF	International Monetary Fund
BIS	Bank for International Settlements
EUREX	Eurex Exchange

Acknowledgements

First and foremost, I would like to sincerely thank my supervisory team, Dr. Mimoza Shabani, Prof. Vassilis Fouskas, Dr. Shazaib Butt, and my former supervisor Dr. Carmela D'Avino. I have been very lucky to have such a fantastic supervisory team. Many people have supported me over my PhD journey, most notably my parents Chris and Sheila, as well as my many colleagues and fellow PhD researchers at the Royal Docks School of Business and Law.

1 Introduction

Financial markets are crucial to all modern economies. Since their smooth functioning is vital to firms and consumers, they attract much scrutiny from academics and non-academics alike.

Since the creation of the world's first electronic trading venue, NASDAQ, in 1971, advancing technology has improved the ability of traders to observe and share information in increasingly short periods of time. This has given rise to a race to trade information into price or correct or close a spread. This temporal 'race' has given competitive traders a 'need for speed' and a strong incentive to reduce the latency of their trading systems. The upper bounds of speed are probably yet to be reached and will be bounded by a mix of profitability and, ultimately, the speed of light. This is the motivation that underpins high-frequency trading (HFT).

The concept of HFT has no exact definition. It is best thought of as referring to algorithmic traders who trade frequently and play latency-based strategies. The activity of high-frequency traders (HF traders) provokes questions of fairness as an increasingly wide asymmetry exists between the technology that HF traders and non-HF traders possess. Furthermore, there are unanswered questions that have been raised in the extant literature with regard to the effects of HFT on marketplace stability and liquidity.

This thesis comprises three chapters that sequentially explore the environment in which HFT operates, an analysis of markets in which HFT operates, and the possible scope for regulation of HFT activities. Chapter Two extends and updates Easley et al.'s (1996) model of market activity to incorporate a latency element that allows for 'trader type' to be incorporated into a model of non-cooperative interaction. This is used to test for a trembling hand stable equilibrium, which in turn is used to test the significance of observed noise within a sample of major trading venues. Here, it is shown that present levels of noise should not pose a risk to traders who are following an information response strategy. Ergo, I show that it cannot presently be concluded that HFT appears to adversely influence market quality through the creation of noise. The third chapter presents a survey of the methods that can be used for identifying HFT activity and seeks to draw out what is known about how HF traders process information in their decision-making process. It also illustrates a method that is based on the application of working price levels and an autoregressive conditional heteroskedasticity (ARCH) estimation to examine individual traders' contributions to the price formation process using ex-post data. This gives proof that regulators and observers are able to assess the price impact individual traders make and retrospective analysis of flagged events is possible. The dataset covers five ForEx pairs over a period of one day and was obtained from Tick Data Market of Paris. This data source is widely available and is not restricted to trading venues and/or regulators. The implication of this model is the ability to, in retrospect, illustrate how individuals affect the market price formation process. This allows for insights into how events developed and provides a framework for researchers to analyse market volatility events.

The final chapter, four, seeks to explore the extent to which existing regulation is effective in maintaining perceived fairness in financial markets that are characterised by the presence of HFT activity. Ideas of fairness vary given it is a very subjective concept, yet it is a frequently cited objection to HFT activity by its critics. Issues of fairness in the context of HFT are centred around access to liquidity and the ability of traders of various types to access the markets and use order types which afford them similar priority in the exchange matching engine. Chapter four proposes a framework of fairness based upon the existing ordoliberal literature and are used to provide a more intellectually acceptable foundation compared to the contribution of Angel and McCabe (2013). This allows for a conceptual analysis of policy and regulatory measures which seek to enable or restrict HFT activity.

A review of present regulatory arrangements illustrates the vulnerabilities of the 'self-regulatory' principle, which essentially places 'Dracula in charge of the blood bank'. Furthermore, large trading entities may also face the problem that supervision of HFT activity is too big to manage. This review is followed by a series of recommendations concerning the monitoring that is undertaken by regulators and the importance of both real-time and ex-post monitoring. In addition, this contribution critiques some of the market structure control mechanisms and illustrates the complexities in restricting some activities in a way that may be seen as unfair.

2 The creation and stability of the signals available to high-frequency traders

2.1 Abstract

Financial markets are subject to increasingly strong latency arbitrage, which has shifted the focus of traders from fundamental analysis to technical analysis for the purpose deploying scalping strategies; these are short time horizon spread capture trading strategies which are typically used in rather liquid markets. As a result, it is necessary to consider the legitimacy of this form of data collection when it is deployed by an algorithmic trader as a basis for position making. This research evaluates the sonority of the chartist signal available over small numbers of ticks and assesses for what time period (and number of ticks) may one chartist signal be expected to provide clarity into a market that can be shown to be efficient over longer durations in which other trader types can act and introduce information and noise into the market. Empirical results indicate that individual assets and exchanges are subject to great idiosyncratic variance with respect to the occurrence of signal and noise in price feeds. As a result, it is possible to show, in the varied sample, that there may be some consistency in price patterns and spacing, which provides 'moments of clarity' for a period of seconds. When considered holistically, these markets may be regarded as efficient in terms of their ability to incorporate all information into prices. This ability elucidates the temporal nature of efficiency and shows that algorithmic traders who engage in scalping need not resort to artificial intelligence (AI), whose system has a sufficient latency advantage for operating within small periods of clarity. This is increasingly possible when HF traders are bounded only by the speed of light.

2.2 Introduction

The aim of this chapter is to analyse the environment in which HF traders operate. Estimates of the proportion of market activity that is attributable to HFT run as high as 73% in the US equities markets (O'Hara, 2014), and this is likely to increase over time. In addition, as HFT 'speeds up' due to the competitive nature of latency-based strategies, the asymmetry between algorithmic and human trading entities has the potential to widen further.

To explore the general market microstructure in which a HFT system operates, existing models of the trading environment are reviewed herein. This chapter also illustrates the role that latency plays in markets and within the non-cooperative market environment, where we assume, at least, partial asymmetry in one trader's ability to sess the actions of another.

To further explore the marketplace and its non-latency characteristics, this study considers information asymmetry in the same terms as in Black's noise trader model (Black, 1986), with the goal of considering the extent to which noise is prevalent in equity markets. This exploration is followed by a discussion of the stability of information-based trading strategies. Here, a Kalman filter is used to produce noise-to-signal ratio values, which are used in a model of market structure to show they have a trembling hand stable Nash equilibrium. This discussion finds that, while HFT may contribute to noise, the observed noise that is present is not significant enough to disrupt an information-based strategy played by a human or algorithmic trader. This finding is novel in the existing literature and addresses a growing concern that HF traders add noise to marketplaces.

2.3 Literature review

HF traders can be characterised as faster versions of human traders (Stenfors and Susai, 2019); however, it may be argued that automated trading systems are simply lines of code that decide what to trade, when, where, and with what order type (Patterson, 2012). These systems use learning and processing methods that include fuzzy logic, genetic algorithms, machine learning, and expert systems. As a result, the ability to process and action information differs between a human trader and an automated trading system (Patterson, 2012). Human traders work in a higher-latency environment and are thus subject to behavioural factors that are not reflected in the automated trading systems.

One of the core aspects in the innovation of such systems is the latency of the system. Such innovations include colocation, 'speed-of-light' transmission technology, and reducing the complexity of the system to ensure information is transformed into strategies and submitted to the exchange (Wang and Zheng, 2015). The objective underpinning reducing latencies is the use of spread capture strategies and a growing latency arbitrage. Arnuk and Saluzzi (2009) describe the reduction of latencies as an 'arms race' to detect an opportunity to arbitrage and place an order that arrives ahead of any other traders in order to capture the first mover's advantage. It is also possible, using the predictive ability of algorithmic systems, to purchase in anticipation of an order and capture the spread; numerical examples are provided in Arnuk and Saluzzi (2009).

Typical spread capture strategies are forms of scalping strategies which are short horizon positions. These are inventory-neutral strategies that seek to capture a spread and exploit the liquidity rebates offered in maker-taker market systems (Patterson, 2012; Bodek, 2013). Such strategies may occur over periods as short as six market updates (often referred to as 'ticks' where the market updates after each trade) in markets which are sufficiently actively traded; in terms of time this could be as fast as 300 milliseconds (about the time a human being needs to blink). Many of these strategies in which predicative ability is used are front-running strategies, where HTFs try to enter the order book ahead of another actor. To this end, the market depth for a scalper is determined by the quantity of the orders behind him (Wang and Zheng, 2015). Another method is the use of sub-penny jumping strategies, where permitted a less than one penny/cent price improvement is offered by a trader to ensure execution before orders are placed at whole prices[1] (Mahmoodzadeh and Gençay, 2017). Manahov (2016) has shown that no benefit arises from this form of HFT activity, as HF traders create many orders that they cancel, thus making it more complex for others to trade. Furthermore, high frequency scalpers have been shown to increase overall market volatility. Manahov (2016) concludes that HF traders do have an advantage over traditional investors and goes so far as suggesting batch auctions may be a preferable alternative. In a batch auction trading system orders are collected over a period of time and then executed simultaneously. This is rather different from continuous trading, where execution takes place at any time, as and when, orders are entered and match. Manahov (2016) argues that this may partly eliminate the advantages associated with latency arbitrage.

[1] For example, if a trader bids 90.1 pence, this will allow him to match before any other trader who has bid 90 pence.

The extent to which this form of activity can be seen as predatory is debateable; however, the activity described is legal and accepted by exchanges. The issue is one of ethical behaviour or fairness, as explained in the study of Angel and McCabe (2013), where the researchers consider that the lack of perceivable fairness due to the existence of latency advantages between traders is a factor in the creation of exchanges that are devoted to specific trader types. The weakness in Angel and McCabe's (2013) argument is that they are unable to fully establish a definition of fairness; however, they quote Shefrin and Statman (1993) in providing an elaborated set of definitions. Broadly, these focus on freedom from coercion, freedom from misrepresentation, equal information, equal processing power, freedom from impulse, efficient prices, and equal bargaining power.

It is possible, given these definitions, to conclude that no market is fair, financial or otherwise. If this argument is accepted, then any financial innovation will be an innovation in an unfair system and shall not, by virtue, be 'making the system unfair' – as journalistic sources have portrayed.

This gives sufficient insight into the rationale behind the use of venues in which random lags are applied to eliminate latency advantages. In addition, this situation has encouraged some traders to avoid lit markets such as the London Stock Exchange (Aquilina and Ysusi, 2019) and instead take advantage of the ability to use dark aggregation ('dark pools'), where it is legal to do so. A dark pool is an electronic venue or mechanism that accumulates non-displayed liquidity and provides matches of bids and offers. A dark pool can take the form of an alternative trading system, internalised order flow, or exchange reserve/hidden orders, and it is designed to minimise costs and market impact while preserving client anonymity. It is noteworthy that the current MiFID regulations do not mandate that large transactions should take place in lit venues;[2] hence, there is a minimum order size for dark aggregation to be possible (Banks, 2009). It is larger orders that benefit most from access to dark liquidity as traders seek to minimise their price impact.

[2] Traditional venues wherein liquidity is visible.

In order to analyse markets with different types of traders, it is necessary to consider in what way they interact with one another. Seminal work has addressed some elements of this issue. Alexander (1961) gives a model that provides a foundation for Black's (1986) noise trader model. Alexander considers a world with two types of traders: firstly, technicians who believe that known facts will influence price in the future, and secondly, fundamentalist traders who seek to gain information as early as possible. It is the latter type who will actively incorporate new information into prices, but the technicians trade on the basis of this information and drive momentum. This situation, by degree of virtue, gives fundamental traders a first mover's advantage, whereas the technician traders are simply participating in a Keynesian beauty contest in which players trade based on anticipation of movements.

An important concept that Alexander (1961) acknowledges is that earnings may be driven by a long-term trend, within which, at short intervals, it would be possible for a fair game to operate. This idea would suggest that, in short periods, markets are efficient, but in the long term, it may be possible to hold profitable strategies that do not necessarily 'beat the market'. Such a strategy is reflective of a long-horizon strategy; however, in liquid financial markets, it is common practice to work in short time periods when adopting spread capture strategies. Therefore, there is a great need to understand how markets are efficient in increasingly shorter periods of time.

Alexander (1961) also points out that professional traders may be reluctant to accept that they are subject to a fair game in the short term; however, much academic evidence supports the view that prices are not predictable. A price can be said to follow a random walk if at any time the change to be expected can be represented by the result of tossing a coin, although not necessarily a 50–50 coin (Alexander, 1961). For this reason, it has been shown that technical analysis has remained popular despite academic evidence suggesting that it has no sound basis (Lui and Chong, 2013).

Niederhoffer and Osborne (1966) make it clear that many people claim to be able to offer explanations for the allegedly random patterns and trends they can see in price charts. These can exist in fair game conditions and Roberts (1959) shows that a random walk can produce the shapes, patterns, and trends that are lauded by chartist traders.

The fair game concept originates in the study of Poincaré (1900), who believed that speculative prices follow random walks, following his analysis of Rentes traded on the Paris Bourse, closely correlated to his predicted distribution reflective of a random walk. Such a position receives validation in the research of Kendall (1953), who used a similar method to compare price with the 29 previous first-differenced lags. He found no evidence of any meaningful relationship between past prices and current price, as random effects 'swamp' any systematic effect that 'may be present'. Kendall (1953) used weekly data from a range of markets in London and Chicago. Kendall did claim to have found only one exception when writing in 1933; however, this has been shown to be a data handling error. Ergo, Kendall's work supports the notion of the fair game. Such an analysis in a shorter time frame could have yielded different results, according to Alexander (1961). In addition, later work by Fama and Malkiel (1970) goes further with the random walk hypothesis and notions of incorporation of information into prices and degrees of efficiency. A clarification here is necessary as these sources are discussing informational efficiency, whereas literature from the game theory field uses the term in connection with Pareto optimality (Shefrin and Statman, 1993). More recent work has sought to test the applicability of these ideas in very short timer periods using more modern means such as machine learning and neural networks, which better reflect human decision making processes and herding activities. Research undertaken by Fischer and Krauss (2018) illustrates that there is a pattern between volatility and frequency of reversals and the longer term return of the asset.

A further model that aims to show interaction between two trader types is the noise trader model. Black (1986), in his noise trader model, hypothesises two types of traders. The first type is smart traders, who react to information and are considered rational in their decisions; these traders should know the correct value of an asset. The second trader type adds noise through poor use of information and thus pushes price away from the rational price. Noise within prices is separate from movement in prices, which is caused during information incorporation.

Investors make their decisions subject to their individual optimisation problems based upon their individual information sets (as is consistent with the two-pillar approach to asset pricing). However, a distinction must be made in terms of information sets comprising both factual knowledge and what is believed to be knowledge but is in fact derived from only noise. This is a fair illustration of the black box of financial markets.

This is a proposition that makes noise seem undesirable; however, Black (1986) argues that noise is crucial for markets to function and even for liquid markets to be viable tools for investment. If a market were devoid of noise, a fully rational trader would not be able to achieve a return. Although, according to Alexander (1961), it may be possible to hold for long periods in order to achieve a return. However, noise prevents the observation of market prices from being perfectly factual observations, and this, in turn, prevents accurate estimations of the future returns of an asset or a portfolio, which prevents an accurate assessment of the current (intrinsic) value (the discounted value of the future expected cash flows). Hence, the existence of noise can reduce the viability of an investment.

If all trading activities did not involve noise, markets would contain much less activity in terms of volume, as investors would hold assets and only trade when necessary in order to change their exposure to market risk. In such a world, rational investors would most likely trade in mutual funds or in portfolios rather than in individual securities. The distinction here is that, without noise, it would not be so necessary to maintain an element of liquidity when investing, liquidity being the ability to convert an asset back to its expected cash value (Hicks, 1962).

Easley et al.'s (1996) model considers signal as given by nature on a daily basis and that traders react to this phenomenon by submitting orders based upon the signal. This model is based on NYSE data, allowing for the estimation of probabilities of informed trading using estimation based on Bayes' rule. By ranking stocks according to trading volume, it is possible to conclude that more heavily traded stocks are subject to more information events and are also subject to a greater degree of uninformed trading. It is likely that no single explanation can address this issue; however, some factors will certainly be behavioural. It seems that two-player models in continuous time have yet to be considered. In such games, it would be necessary to consider the ability of one player to impact the prices available to another player. Price setting is conditional on market liquidity and the trader types within it (Johannsen, 2017).

Prices are set in environments that can be elucidated according to game theory. However, it must be acknowledged that price discovery is different in low-latency environments and environments in which liquidity differs. For instance, liquidity differs in availability based on trader type, depending on what trader type serves as the designated market maker (DMM) (Stenfors and Susai, 2019). According to Johannsen (2017), information arrives and is processed at different speeds, creating the opportunity for arbitrage between exchanges. This is known as toxic arbitrage,[3] and, while serving as the basis of a strategy for an HFT, also reflects how information moves between venues and may lead to overreactions and further revisions, thus introducing trembling prices and potentially noisy (resonant) amplification. It is suggested that toxic arbitrage has greater prevalence in low-liquidity markets; however, much of the existing work has examined equity and ForEx markets. The assumption is that arbitragers act in at least two markets and are willing and able to face the cost of doing so and that they are able to act within the time period during which the information remains asymmetric between the two markets. Actions depend on the bid-ask spreads in each market compared to the magnitude of the price change in the more informed market. Here, no exotic order types are needed (Johannsen, 2017).

One must be cautious about information use in arbitrage strategies with a latency dimension. 'High frequency traders do not care if information is accurate' (Foucault et al., 2016, p. 335). This is because algorithmic trading systems (ATSs) typically have a notably short-term objective and sess information differently to long-horizon traders. It is also the case that HF traders are more likely to use strategies based on technical analysis; to some degree, this is due to the comparatively slow evolution of fundamental indicators.

Following Johannsen (2017), it can be shown that it is possible for a well-placed trader (with very low latency) to try and trade ahead of news. This allows HF traders to aggressively pre-empt price movements and contribute significantly to the trading volume that arises due to an information event. Here, the signal needs to be low latency and high frequency. The price feed is the usual means, but developments in AI and

[3] This term seems to originate in the study of Foucault, Hombert, and Roşu (2016).

progress with machine-readable text allow traders to work with a much larger information set when trading ahead of the market (Pang et al., 2002). A key issue here is that, in the short term, liquidity providers are exposed to losses due to the activity of the aggressive HF traders when the toxic arbitrage strategy is deployed against them (Foucault et al., 2016).

Previous work has focused on the ForEx market, as it is characterised by a mixture of trader types. These also remain an area of interest, as the markets are free from circuit breakers; therefore, mass withdrawal of liquidity is a concern, as this can amplify the effect of a shock to the market, especially where the role of the specialist is held by an ATS (of the high-frequency type) acting as the designated market maker, whose role is to maintain the provision of liquidity (Stenfors and Susai, 2019).

It is also worthwhile to consider the variation in trader types, which is based upon the information sets they use, their decision-making processes, and their reaction times. Stenfors and Sasau (2019) examine the ability of different types of traders to interact within the same market.

In considering liquidity it is important to consider price, volume, and speed in the analysis. Prices determine how viable it is to change a position (bid-ask spread),[4] and volume shifts price[5] and speed determine the time that is needed to recover from an exogenous impact. HF traders operate in terms of milliseconds, microseconds, or nanoseconds, compared to the several seconds that a human (via a computer) takes to take the same action. However, an additional difference is the use of heuristics or even the concept of a ‘sense of duty’. Empirical work that uses data sources from Electronic Brokering Services suggests that liquidity is provided in response to a limit order (as expected); however, this liquidity begins to be withdrawn after only 0.2 seconds – between one and five seconds after the order the book has reverted to its original position. Ergo, this liquidity would be accessible to a human trader (Stenfors and Susai, 2019).

[4] Some markets lock when the spread is zero.
[5] Unless the volume is very small, the orders are dark and/or well shredded.

Barclay and Hendershott (2003) consider at what time information enters markets and how those shocks occur. They consider trading after hours (as this is possible on the NASDAQ exchange). After-hours trading is undertaken at the discretion of the individual broker; however, this decision must made with the best execution obligations in mind. It is likely that informational asymmetry varies across periods of the day; for example, there should be less asymmetry in the post-close than in the pre-open periods. As a result, the probability of a trade being an informed trade also varies throughout the day. Ergo, it is possible that noise similarly varies and that markets handle information differently, as outlined in Alexander (1961) and discussed previously.

The trading day itself yields the greatest amount of price discovery; however, the pre-open and post-close periods are hosts to a greater degree of informed trading. Private information is disclosed in the pre-open period, and this starts the process of declining information asymmetry over the day. The conclusion is that most activity that shifts price occurs at night. This is consistent with the well-known phenomenon of overnight risk. The present author adds that, even in an exchange without the extended window, there still should be declining asymmetry over the course of the day, and, as a result, some traders rationalise the strategy of observing and waiting.

It follows that models that incorporate a single daily flow of information are not realistic. For example, Easley et al. (1996) consider signal strength to be set by nature daily. This is a simplification that allows the model to be more widely applicable to financial markets rather than just equity markets. Such trade-offs in specification are discussed by Sutton (1997).

Wilson (1985) considers agents' strategic behaviour when trying to develop a complete analysis of financial markets. In modelling financial markets, Wilson considers number, endowments, preferences, and information, which are set within a given regulatory structure to allow for an explanation of price formation. For each layer, their strategy set is formulated as a series of contingencies (whenever possible, the probabilities should be attached).

In 1985, game theory was advancing sufficiently, allowing for an application to pricing so long as the games are constructed with explicit trading rules (omitting the possibility of a market mediated by specialists[6]). Trading rules are specifications that determine possible actions, thus limiting the possible outcomes (Wilson, 1985).

In Foucault et al.'s (2016) model, it is possible to suppress the concept of private information, so it is obviously possible for a model to consider this element. When we take private information to exist for each party in a game, the trader with the greatest information set likely receives the largest gain from a trade, although trading can reveal the valuation/information that they know. Exposing this knowledge could reduce this trader's gains from the trade. This is the logic behind dark aggregation, so it needs to be a component of the strategy set. In some cases, waiting can also reveal a signal to other players. This is particularly easy to conceptualise if one thinks of a Dutch flower auction (Wilson, 1985), in which the price moves from high to low and the first person to bid wins the auction, which indicates the winners maximum willingness to pay to all who observe the auction.

Allen and Morris (1998) express a need to adopt von Neumann and Morgenstern's (2007) approach to choice under uncertainty. This expected utility approach is seminal in game theory and asset pricing. While the approach restricts what we can conclude about irrationality, it is possible to confirm optimality through the use of a trembling hand Nash equilibrium which allows for the testing of a Nash Equilibria where a probability of incorrect strategy selection is accounted for (Samuelson and Zhang, 1992). The present author doubts that a direct measurement can be made of the trembling hand in this research, although it may be possible in an experimental economics setting, which is beyond the scope of this study.

The efficient market hypothesis (EMH) holds that the only factors that may be deterministic of price are those that affect risk, as these alone affect price. Caution may be virtuous here, as a very literal interpretation of EMH could lead to the view that information is wholly symmetric between traders. Current models can be expanded by including transaction costs (a stated restriction in the research of Black

[6] This thesis considers markets with a market maker and thus cannot adapt Wilson's conclusions to modern conditions.

[1986]), which are limits to arbitrage and thus may be limits to the incorporation of all information into prices.

Allen and Morris (1998) cite research that assumes that players can learn from prices, but each player holds the belief that they do not influence prices. The present author finds that this concept is incorrect, as supported by Sutton (1997). It is more realistic to argue that actors do know they pose a price impact, as this is the basis of shredding strategies and dark aggregation. Furthermore, this situation is partly why traders are cautious never to cross orders.

The ability to process information is developed as part of a learning process, and here, past events and actions are considered to have the ability to influence the use of the strategy set that a player has at their disposal in the future. It is necessary to consider the speed at which the player learns via the principles of reinforcement, and this can be linked to the development of heuristics, which enable the use of system-one thinking which is dependent upon previously learnt heuristic shortcuts rather than active thought (Bereby-Meyer and Roth, 2006).

The speed of learning depends on the degree of noise in the payoffs within the game (Bereby-Meyer and Roth, 2006). In plain terms, this means that the more predictable the payoff structure, the quicker a player will develop 'learned cooperation'. In the case that the payoff structure is noisy, the reinforcement is only partial. Relevant literature suggests that behaviour can develop faster in an environment of deterministic games; however, trading is always a probabilistic environment. This is largely due to asset prices generally being held to follow a random walk, but it is also due to asymmetry of information (prices move, but an individual will not always know why). In the probabilistic game, the speed of learning then depends upon the presence of a dominant strategy. In financial markets, the unpredictability of other players erodes the continual presence of a dominant strategy (for example, the one-shot dominant strategy in a prisoner's dilemma game, where the game is not repeated).

Some evidence does challenge the orthodoxy of the random walk hypothesis (Kemp and Reid, 1971). Jegadeesh and Titman (1993) illustrate that strategies which buy previously high preforming stocks and sell previously poor performing stocks earn a super-normal return over a period of 3 to 12 months. This indicates that a behavioural element may also explain price, not just random delivery of information.

As evidenced, the payoffs of a game are determined by prices; henceforth, the payoffs are not always predictable, given than it is common to form strategies using von Neumann and Morgenstern preferences, which take the expectations of payoffs in the future – conservatively (Allen and Morris, 1998). Here, it may be argued that increasing latency helps improve the accuracy of predictions.

Bereby-Meyer and Roth (2006) discuss the possibility that players can reward and punish each other in future games. Although this is valid in some games, in the case of financial trading, the anonymity of the games and the fact that a player never knows with whom their order has crossed make Bereby-Meyer and Roth's proposition difficult. The only exception is when a person crosses an order against themselves (HFT can do this; however, it is most undesirous [Wang and Zheng, 2015]). Bereby-Meyer and Roth (2006) conducted behavioural experiments to arrive at their conclusions; however, the present author cautions that these are based on single-period prisoner's dilemma games and repeated prisoner's dilemma games that do not perfectly reflect financial markets. Their research does not attempt to incorporate a scenario where multiple players match each other randomly. For the purposes of this thesis, the research of Bereby-Meyer and Roth (2006) is useful; however, caution must be taken in applying it. While learning depends on noise and trading is indeed a continuous game, the models cannot be replicated in an exact sense.

2.3.1 Models of the trading environment

A noteworthy model of trading is found in the study of Easley et al. (1996). This research introduced a model for the analysis of continuous time. Here, trades are considered to be either informed or uninformed; the distinction in this case is that only informed traders react to news events and create strategies that profit based on the first mover's advantage in the price update. Easley et al. (1996) accept that such is the crux of an efficient market; as a result, their findings are compatible with other seminal material.

It is helpful to first introduce the model and then discuss some of the findings concerning the application of the model to a dataset.

Here, investors of two types are considering a single risky asset over a period of t = 1, …, T trading days. Each day takes one of three forms, which determines the information flow; this is derived from nature and thus is exogenous to the model.

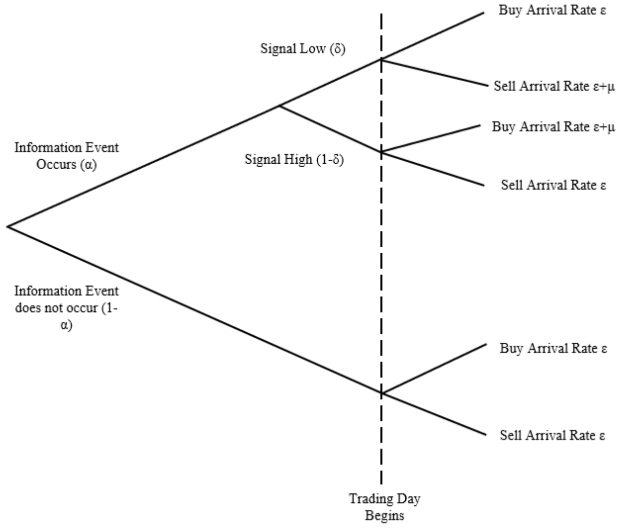


Figure 2.1 Extended form of the model of Easley et al. (1996).

The order flow depends on the assigned state of nature that occurs with the probabilities α and γ, as shown in the extended form in Figure 2.1. Information yields the order flow measured by ε and μ, where ε is the rate of uninformed trading and μ is the rate of informed trading.

In the original diagram in Easley et al.'s (1996) manuscript, a dotted line appears; however, it is not the case that asymmetry is held, as informed traders need to be able to sess the state of nature. It is likely the case that uninformed traders can also sess; however, they are insensitive to the information.

In this model, the periods considered are day by day; however, it should be possible to apply the same model to shorter periods in order to assess how informed trading may vary over the course of a day. This is a pertinent line of enquiry as it is known that information incorporation varies throughout the trading day as outlined by Barclay and Hendershott (2003).

Using a sample of stocks randomly selected from NYSE, it was possible for Easley et al. (1996) to observe the daily values for parameters α, γ, ε, and μ. These can be used within an application of Bayes' rule to calculate the probabilities of informed trading.

The pairing of conditional probability to produce measurements is possible when the classification of signal is possible; in retrospect, however, this model does not explicitly address the quality of the information or how the traders learn. This is a very useful model, but it is not replicable in terms of the research in this chapter given the limited nature of the data.

The sample in Easley et al. (1996) was ranked by deciles based on trading volume. The conclusion is that more heavily traded stocks are subject to more information events and are also subject to a greater degree of uninformed trading**.** The element of the signal being received from nature will be replicated in further models that aim to introduce a multiple trader environment to the model.

In order to consider a game that has player types related to latency, it is necessary to use a sequential game structure whereby one player has the first mover's advantage and the following players can observe price impact – but nothing more. Much literature has been devoted to sequential games, in which it is necessary to develop a mechanism to allow for the price impact of earlier players to impact later players (Kokot, 2004). In business management literature, it is common to set advertising values or R&D costs in a qualitatively high or low sense (Sutton, 1997), which allows for the introduction of latency into models; however, this practice perhaps also necessitates an alteration to the sequential nature of the game (fast vs. slow) to allow for the modelling of front running as an order-changing strategy. Thus, an alteration to the current practice in order to allow for the specification of all contingents within the bounds of the market microstructure is also needed (Wilson, 1985).

2.3.2 Asymmetry in payoff structures

Work evaluating ultimatum games has been used to evaluate fairness and willingness to cooperate or defect from a game. In experimental games, the asymmetry of the payoff structure is varied throughout the repetition of the games. This alters the sequential structure, as in a purely sequential ultimatum game, player one would afford player two the minimum positive amount possible; however, when the asymmetry is relaxed, a notion of fairness is introduced (requiring player one to make an offer that is acceptable to player two, as player two is able to punish player one[7]) (Kagel et al., 1996).

In the case that both players are fully informed of each other's actions and payoffs, there is a conflict between the typical strong monotonic reactions (as in the von Neumann–Morgenstern utility theorem) and in 'fairness' normalisation. Thus, heuristics can present a challenge to the game theory models in the pure terms that Allen and Morris use to describe them.

When, in an ultimatum game, the first player is the only informed player, Kagel et al. (1996) found (on average) a three-to-one income split in favour of the first player. Probit regression models offer empirical support for this finding, as the researchers found that the only significant factor in rejections was the number of units offered by player one to player two. When both players are fully informed, the rejection rates were much lower and the payoff was more equal; however, from observing player two's responses in the sample, it was also noted that player two would give the benefit of the doubt when they did not have an information set and would require a greater sense of fairness when they did. Kagel et al. (1996) also provide a comparison between the ultimatum game structure and the Nash bargaining game, the principal difference between the two being that, in the bargaining game, the players are able to bargain and reach an agreement (which is possible within constraints, such as time). With common knowledge, rejection rates are lower in games that allow bargaining and shared information sets.

In the case of asset pricing, it is not possible for all players to have a full information set. Rejections may occur when an order is not matched, and this is a consideration when a player is deciding at what depth to place their order in the order book. However, it is sufficient to conclude that prices are equilibrium outcomes by virtue, regardless of the mechanisms around them. However, they cannot always be Pareto optimal, and the notion of fairness is excluded, which causes a reversion in benefit of the doubt or possible trust in the 'system', which is an entirely human concept (Stenfors and Susai, 2019).

---

[7] And he is also punishing himself.

2.4 Models of the multi-trader environment

This section seeks to illustrate the concepts in the existing literature in Models A and B, before outlining a larger Model C in order to discuss market microstructure by applying a payoff structure.

Easley et al. (1996) consider the continuous model structure and states of the world determined by nature, and the structures in this thesis remain consistent with these ideas. However, the proposed models are of the multiplayer type and are characterised by the existence of partial asymmetry to allow players to see that another has traded, but they do not understand why or what strategy was used.

2.4.1 **Model A**

Model A establishes that the first mover's advantage can be taken as exogenous (to simplify the model). Here, each player is given the same strategy sets in a sequential game; however, the asymmetry applied is only partial (as shown in Figure 2.2) to allow the second player to sess if the other has traded or not, but there is no more information than that. A key assumption here is that all orders are market orders, so the queue to execution is based upon the time the order is received only.

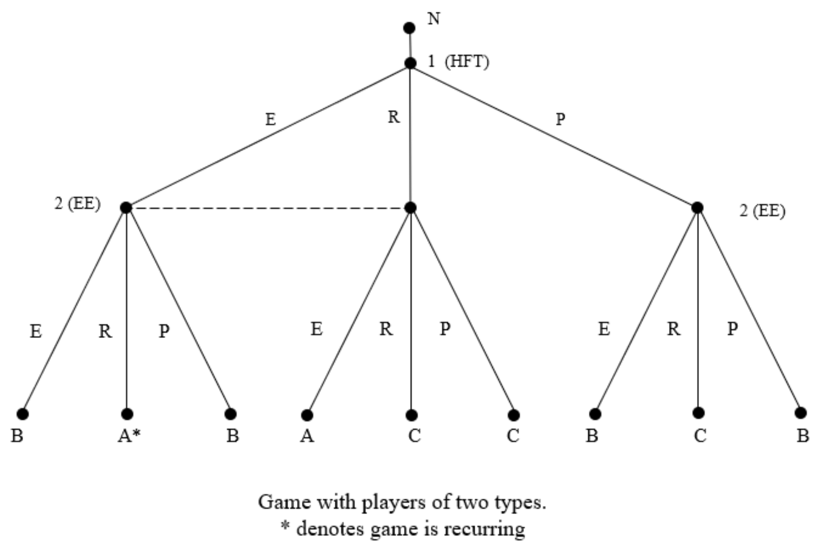


Figure 2.2 Extended form of a game with two player types. In this game, the payoffs are theoretical and are denoted by letters A, B, and C.

Possible strategy set:

E – Eliminate

R – React

P – Pass

Possible outcomes/payoffs:

A – Cross-information removal

B – No incorporation into price

C – Information traded into price

In this game, the HF trader is player one and moves first. It is assumed that the HF trader has a latency advantage and can respond faster than the slower (EE[8]) trader to the signal provided by nature (N). Here, the HF trader is modelled as having only three available strategies:

1. React – The signal is identified as valid information, which will be traded into price using an appropriate strategy.
2. Eliminate – The signal is treated as noise, and the trader attempts to eliminate the noise through arbitrage.
3. Pass – The trader decides that no action is necessary or that the opportunity is not sufficiently profitable to exploit.

The asymmetry between players is only applied partially as the second (slower) player cannot determine what classification the HF trader has applied. They only see that a trade has occurred through the market data feed, which may or may not also indicate a price impact. However, it is possible to see that a lower-latency trader has not traded; hence, the asymmetry is only applied partially.

In this game, it is theorised that information can be incorporated (Payoff C) or that signal is not traded into price (Payoff B). It is considered possible for information to either be cross-removed or to be noisily amplified (Payoff A). The existence of the partial asymmetry characterises the 'black box' of finance and reflects the ability of market stakeholders to observe changes in the asset price yet not understand the rationale behind these changes.

[8] EE generally is taken to mean 'everyone else', or everyone who is not an HFT.

When there is an overreaction to signal the exogenous impact will be amplified by the creation of noise which will be arbitraged away be rational investors. This 'noisy amplification' characterises the need for traders of both types to be able to make a distinction between signal and noise when using market data feeds in low latency environments.

2.4.2 **Model B**

This second model extends Model A to incorporate the opportunity to front run the player with the first mover's advantage. For this situation to occur, the model must consider two HF traders of varying latencies (the exact latency is not important here; however, in Model C, the payoff structure can incorporate this parameter).

It is possible to illustrate markets wherein two players of the same type are interacting with each other. In this model, two HF traders are shown, one with a first mover's advantage over the other.

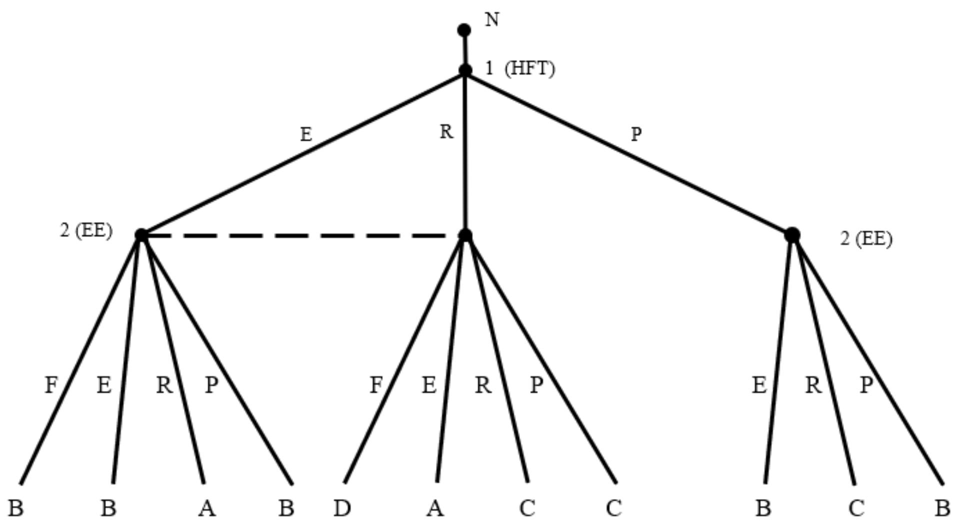


Figure 2.3 Model B in the extended form.

Possible strategy set:

E – Eliminate

R – React

P – Pass

F – Front Run

Possible outcomes/payoffs:

A – Cross-information removal

B – No incorporation into price

C – Information traded into price

D – Incorporation [with welfare loss]

A new strategy becomes available: the option to attempt to front run. The second trader attempts to submit an order ahead of the first trader. In this instance, the strategy is only successful if latency is sufficient to place a superior order type into the market before player one's order is actioned by the exchange matching engine. The availability of this strategy set allows for the introduction of welfare loss (Payoff D), as this allows the same result to be achieved as in Payoff C – but via a greater number of trades than was needed to achieve the same result as in Payoff C. Ergo, it can be thought of as a welfare loss or an erosion of payoff.

If player two was able to front run against player one, player two's order would reach the matching engine first and remove the effect of player one's first mover's advantage. In this case, it has been possible for the sequential game to play out in a non-sequential way. Thus, a clear distinction is now needed when attempting to model the markets. The order of the game's sequence is set by the information flow from the state of nature to the traders, after they have elected strategies that are in response to the state of nature in the game. The order of the flow to the matching engine of the exchange cannot be captured in the extended form and thus is only reflected in the payoffs.

There is a degree of welfare loss; however, game theory models are better suited to the assessment of efficiency. As a result, no comment on the fairness of this strategy can be made. This is consistent with the earlier discussion of 'fairness' based on the findings of Angel and McCabe (2013).

2.4.3 **Model C**

Models A and B are simplified explanations of and introductions to the concepts of the market microstructure and clarify the need to consider markets as parital asymmetry games. As the nature of partial asymmetry and of ordering have been clarified in simple illustrations, they can be further simplified in Model C, where a more elaborate payoff struture can be seen.

This is a complex model; however, it is limited, as it consists of only two traders, as it is impractical to consider more than two traders and would be overwhelming to show the work of the matching engine and the whole depth of the market.

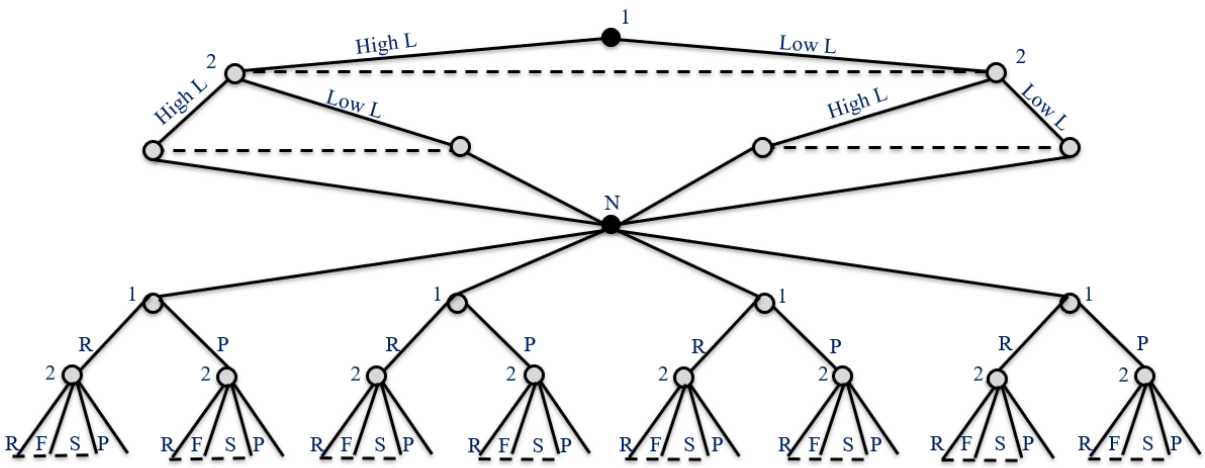


Figure 2.4 Extended form of Model C. L denotes latency.

This model allows for signal polarity, as in Easley et al. (1996). It determines the long/short direction the traders will take and thus affect the payoffs under the maker-taker pricing structure, whereby those who provide liquidity at a point in time receive a partial rebate upon their exchange transaction fees. Otherwise, the game plays out in the same way, regardless of the long/short position.

This model adds a second strategy to alter the order of execution. Sup-penny jumping is considered in addition to the front-running strategy in order to allow the player without the first mover's advantage to reach the matching engine first. This practice incorporates the concepts highlighted in the study of Mahmoodzadeh and Gençay (2017).

With some assumptions, it is possible to show the theoretical payoffs to players in an extended game wherein players begin by investing in their latency to allow them to make market orders. Higher-latency players face greater investment costs (C), but can face lower trade costs (consistent with the findings of research such as that of Wang and Zheng [2015]). As it is difficult to apply real data as latency data and the costs to reduce latency are not available, the following is an illustration of the payoffs to each player.

Firstly, setting latency gives a system cost per trade (S):

$$S = \frac{C}{\sum_{i=t-n}^{t} Vol + \sum_{t+n}^{t} E[Vol]}$$

Here, the total system cost (C) is spread over the number of trades made in the system's lifetime (past and future); as a result, this can only be theoretical measure based upon the use of an average cost per trade. However, traders who seek latency advantage evolve their systems at a faster rate in the 'arms race'; hence, the denominator should be smaller, but the aggressiveness of the trading strategy is also a factor.

Assume trade costs (z) are inversely related to *S*:

$$Z = \frac{1}{S}$$

In the maker-taker pricing structure, those who provide liquidity receive rebate *r* as the incentive to provide liquidity in the market. Liquidity-making traders receive, in addition to their transaction cost:

$$z = \frac{1}{S} - \left(\left(\frac{\mathbf{1}}{\boldsymbol{S}}\right) \cdot \boldsymbol{r}\right)$$

where *r* is the rebate value (e.g. 0.4).

Payoffs are not calculated based on the order shown but rather on the order in which orders are received in the matching engine (as the strategies of player two allow them to jump in front of player one).

The first to execution receive the value of the spread *x* multiplied by the proportion of the spread that they close *s* – less their transaction cost.

Equation 1

$$\pi = (x \cdot s) - z$$

The second receives:

Equation 2

$$\pi = (x \cdot (1 - s)) - z$$

Here, $s + (1 - s) = 1$, so we assume that all of the opportunity is exploited by the two players.

If a player has chosen to play the pass strategy, they receive a payoff that is equal to zero, and if the other player has chosen any strategy, they will earn the payoff shown in Equation 1. It is not possible to earn a negative payoff, as system costs are divided by trading volume in this model to give a cost per trade: s.

In the case of front running, it is necessary to extend the structure to cover the cost of running a system that is capable of front running. The model already covers the costs associated with latency; however, the additional information sessing and processing required to front run effectively must be borne on a per front-running trade basis and is denoted by f, which will always bear a strictly positive value: $f > 0$. Hence, Equation 3 depicts the payoff to a front runner:

Equation 3

$$\pi = (x \cdot s) - (z + f)$$

A player against whom a front-running strategy has been performed receives Payoff 2. The value of s is likely to be rather high in Equation 3; hence, the profit in Equation 2 is likely close to zero.

Furthermore, it is necessary to make a small adjustment for the player who uses a sub-penny jumping strategy, which narrows the spread captured. The Securities and Exchange Commission's Rule 612 has restricted minimum price improvements in order to limit this strategy (Rule 612 [Minimum Pricing Increment] of Regulation NMS, 2005). For stocks priced below \$1, the minimum improvement is \$0.0001, or the minimum improvement is \$0.01. As a result, the minimum improvement should be subtracted from x, where x is denoted in the exchange currency (e.g. USD). This reduces the profit, as shown in Equation 4; however, it may allow for a greater value of s.

Equation 4

$$\pi = ((x - 0.0001) \cdot s) - (z)$$

It is necessary to clarify which trader recovers the first and second payoffs using all strategy combinations. Here, the react and eliminate strategies are merged in order to simplify the game and remove a distinction that the game cannot capture. In very short time periods, the two actions are the same.

Table 2.1 Model C in the normal form – payoffs denote first and second payoffs, as defined in this chapter.

N: Signal is **HIGH**

1 / 2	L, R	L, F	L, S	L, P	H, R	H, F	H, S	H, P
L, R	1, 2	2, 3	2, 4	1, 0	1, 2	2, 3	2, 4	1, 0
L, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0
H, R	2, 1	3, 2	4, 2	1, 0	1, 2	2, 3	2, 4	1, 0
H, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0

N: Signal is **LOW**

1 / 2	L, R	L, F	L, S	L, P	H, R	H, F	H, S	H, P
L, R	1, 2	2, 3	2, 4	1, 0	1, 2	2, 3	2, 4	1, 0
L, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0
H, R	2, 1	3, 2	4, 2	1, 0	1, 2	2, 3	2, 4	1, 0
H, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0

In Table 2.1 above, the cells shaded in grey denote an impractical strategy: If player one passes, it is not practical for an order-changing strategy to be used by player two. This would be a valid strategy, but it is inefficient, as these strategies require greater costs in order to trade.

It is now necessary to establish solutions to the game environment based on the five possible outcomes that comprise the four equations in this section and the null payoff.

This model can be solved by elimination of dominated strategies to four pure strategy Nash equilibria.[9] Appendix 2.2 outlines the solution using the mutual best responses method, which is a necessary proof, as it is possible that a solution by means of elimination will not identify all Nash equilibria.

The conclusions of a solution by elimination in terms of iterated dominance follow:

1. For both players, action strictly dominates inaction (all *P* strategies are eliminated).

2. Front running is strictly dominated by sub-penny jumping for player two in all cases.

3. For a slow trader, sub-penny jumping strictly dominates reaction strategies.

4. For a low-latency player, reaction weakly dominates sub-penny jumping.

The two Nash equilibria are (*HR, LR*) and (*LR, HS*), leading to the following conclusions:

1. Traders without the first mover's advantage should use the sub-penny jumping strategy to gain first execution.

2. Where the trader with the lowest latency does not have the first mover's advantage, they may gain it with the aid of the sub-penny jumping strategy.

2.5 The Kalman filter as an indicator of signal strength

Before addressing the strategy set of an individual trader, it is first necessary to establish that a variance in prices reflects information that is not known to every participant in the market; however, the prices also contain noise (variation without a basis). Ergo, if it is assumed that prices reflect information and that markets are at least semi-strong in their efficiency, it is necessary to establish that there is not much noise in the market. This will allow for an elucidation of the actions that players can deploy.

---

[9] Strategies from which, *ceteris paribus*, neither player will be willing to deviate.

2.5.1 Dataset

The sample is a random selection of stocks over four exchanges: London Stock Exchange (LSE), Paris Stock Exchange (PAR), Hong Kong Stock Exchange (HKEX), and New York Stock Exchange (NYSE). Stocks were selected at random. A total of 15 shares were examined over a period of 10 days, between 07 January 2019 and 18 January 2019. In total the dataset contains 1,220,299 tick observations. This sample can display effects that are specific to the share and to the exchange. It is also possible to capture the effect of the trading volume on signal purity. Data is taken from Bloomberg LP (2019).

Table 2.2 A summary of the sample

Stock	Ticker	Exchange (Index)
AIA Group Ltd	1129 HK	Hong Kong (Hang Seng)
Apple Inc	AAPL UW	New York (S&P)
BAE Systems	BAE LN	London (FTSE 100)
China Unicom Hong Kong Ltd	762 HK	Hong Kong (Hang Seng)
Geely Automobile Holdings Ltd	175 HK	Hong Kong (Hang Seng)
HSBC Holdings Ltd	HSBA LN	London (FTSE 100)
Klepierre SA	LI FP	Paris
Metropole Television SA	MMY FP	Paris
Nielsen Holdings PLC	NLSN UN	New York (S&P)
Royal Mail Group	RMG LN	London (FTSE 250)
Salesforce.com Inc	CRM UN	New York (S&P)
Seven Trent Water	SVT LN	London (FTSE 250)
Tullow Oil	TLW LN	London (FTSE 250)
Vicat SA	VCT FP	Paris
Yum! Brands Inc	YUM UN	New York (S&P)

For each of the 15 stocks shown in Table 2.1, the price and the trading volume for a period of 10 trading days were captured. Thus, it is possible to calculate a daily signal-to-noise ratio, daily volume, and the number of turns in the price feed. This sample size is sufficient for illustrating the data characteristics and allows for an understanding of the relationships between the variables that have been calculated.

Reversals were identified within the price feeds of each asset. Reversals mark a change in the direction of the movement of the price. It is possible to observe a change of direction in as little as three ticks of the market; however, a fourth tick is needed to confirm that there is an onward progression of the price path. This 'extra' tick is required in order to allow for old resistance to become new support and for old support to become new resistance. The origins of this logic lie in Dow theory, the earliest available source for which is Ormerod (1939). When price turns upwards, this is a 'bull' turn, and when price turns downwards, this is a 'bear' turn.

As the identification of patterns is rather a topic of technical analysis, it is noteworthy that detecting a reversal over four ticks presents a very short-term view of the market.

A bullish reversal (price was falling but is now rising):

$P_{-3} > P_{-2}$, $P_{-1} > P_{-2}$, $P_0 > P_{-1}$ and $P_0 > P_{-3}$

A bearish reversal:

$P_{-3} < P_{-2}$, $P_{-1} < P_{-2}$, $P_0 < P_{-1}$ and $P_0 < P_{-3}$

A turn is visible by examining the latest price and three previous lags, which can either be bull or bear turns. A bull turn occurs when price ceases to decrease and begins to rise, and it is necessary to confirm the shape by asserting that the price at time t is greater than the price at time t-4. In total the data set contains 140, 105 of these reversals.

2.5.2 The Kalman filter

The Kalman filter is a short memory process which can be used to indirectly measure variables which are not directly observed. In this application the Kalman filter is used to assist in determining the relative proportion of noise to signal in financial markets when we consider them on a tick-by-tick, intraday basis. This allows for the use of a proxy measure of noise, to be used to consider how effective strategy rational formation can be as a trading strategy in the current financial markets.

Using a small sample of highly traded currency stocks, each were subjected to a Kalman filter to separate signal from noise based on the filter's algorithmic process: The higher the ratio, the greater the amount of signal in the feed.

The Kalman filter is a state space model that has been used extensively in computing and engineering, and it is now being used as an econometric tool in finance due to its ability to use one-step-ahead prediction to smooth a series and capture the error (difference). At first glance, this filter may be confused with a Hoderick-Prescott filter; however, the Hoderick-Prescott filter is designed to de-trend (typically macroeconomic) data. Here, the Kalman filter does not remove the trend. It is possible to add a cyclic element to the Kalman filter; however, this is not necessary here. Hence, the general specification is applied as outlined below.

The system of equations is as follows:

Equation 5

$$y_t = \tau_t + \varepsilon_t$$

ε_t represents the minimum mean square estimate of the observed y_t at time t given all previous values of t, which captures the innovation in this step. τ_t is the trend element that comprises all previous values of y_t.

Equation 6

$$\tau_t = \tau_{t-1} + v_t$$

v_t captures the update in the trend (the state series). Both ε_t and v_t are taken as iid $N(0, \sigma^2)$. In addition, $E(\varepsilon_t v_t) = 0$. Y_t is the independent variable, and τ_t is the trend element.

The state space model was manually specified using the syntax provided in Appendix 2.1. This model is derived from the research of Bossche (2011); however, the setting of the system's starting values is influenced by the discussion in the study of Rummel (2015) concerning the complexity of setting starting values for the two coefficients used in the estimation of the filter parameters. The traditional approach is to adopt a lengthy trial-and-error approach to achieve the fastest convergence (with the fewest iterations); however, this can be considered data mining. To avoid this potential issue, it was decided to set the starting value for $c(1)$ and $c(2)$ equal to the measurement error of the data, expressed as an first-order autoregressive process ($AR(1)$).

It is important to verify that the data is not best expressed as an autoregressive moving average process of order one ($ARMA(1, 1)$), as this may cause false estimation and the starting values would not be efficient. The standard method is the correlogram, which shows autocorrelation and partial correlation. It is therefore possible to verify that the $AR(1)$ component applies in all cases; however, it is more difficult to 'eyeball' the $MA(1)$ processes, and, in some cases, these cannot be ruled out.

Equation 7

$$\boldsymbol{q = \frac{\sigma_v^2}{\sigma_e^2}}$$

The calculation takes the error terms from both the signal and the state equations to produce a signal-to-noise ratio.

2.5.3 Signal-to-noise ratios

Table 2.3 Summary of the dataset

Code_date	Signal-to-noise ratio	Daily volume	Code_date	Signal-to-noise ratio	Daily volume
175_8	1.00072729	29543	NLSN_7	6.84356E+11	1693
175_9	0.46506877	41305	NLSN_8	1.21141E+19	1666
175_10	0.484136256	17845	NLSN_9	7.6305E+160	1893
175_11	0.392047604	11153	NLSN_10	1.83576E+13	1749
175_14	0.560829544	8268	NLSN_11	1.87535E+87	2558
175_15	0.721658407	10035	NLSN_14	1.60558E+41	1677
175_16	0.914329003	11503	NLSN_15	1.0472E+17	2446
175_17	1.000349732	16769	NLSN_16	2.4273E+63	2243
175_18	0.610425097	12285	NLSN_17	5.1416E+32	1816
762_7	0.416571161	3879	NLSN_18	6.49901E+40	1393
762_8	0.334160229	5123	VCT_7	2.9304E+17	145
762_9	0.341219385	9594	VCT_8	6.764E+147	267
762_11	0.268606302	4895	VCT_9	6.2788703	468
762_14	0.311651542	3586	VCT_10	30.86948306	293
762_15	0.433900629	5217	VCT_11	3.775405691	346
762_16	3.645475124	3232	VCT_14	2.15564E+11	208
762_17	1.253767541	3371	VCT_15	6.614109234	443
762_18	1.656115745	3036	VCT_16	96.14018144	294
1299_8	0.279228189	11954	VCT_17	3.32892E+13	442
1299_9	1.000311441	13824	VCT_18	1.17651E+15	338
1299_10	0.639724587	14423	YUM_7	1.25292E+14	1962
1299_11	0.532695208	12327	YUM_8	6.78574E+30	2174
1299_14	0.336042935	10870	YUM_9	1.8331E+39	1509
1299_15	0.539293203	14371	YUM_10	7.86282E+12	1565
1299_16	0.727404072	15174	YUM_11	1.000702522	1559
1299_17	0.574597357	13200	YUM_14	52.88284482	908
1299_18	0.47087493	11261	YUM_15	1074.103165	1120

AAPL_7	3.60697E+26	55124	YUM_16	7.04102E+91	1052
AAPL_8	5.11619E+20	42293	YUM_17	38.88522004	1633
AAPL_9	6.7257E+118	44497	YUM_18	1.89573E+62	859
AAPL_10	1.44144E+76	35936	BAE_7	1.000017299	7902
AAPL_11	5.64379E+85	29691	BAE_8	1.00000014	11160
AAPL_15	65765157874	29747	BAE_9	1.000001082	12942
AAPL_16	4.8843E+231	31275	BAE_10	0.999997536	8414
AAPL_17	1.06184E+64	30306	BAE_11	1.000015428	7772
CRM_7	1.61452E+18	7053	BAE_14	1.000010568	7128
CRM_8	1.001521973	6623	BAE_15	0.999995116	3759
CRM_9	8.1967E+17	3773	BAE_16	0.999999999	3650
CRM_10	1.05111E+78	3613	BAE_17	1.000000639	6988
CRM_11	4.45062E+13	3058	BAE_18	1.000000552	5335
CRM_14	482.965795	2456	HSBA_7	1.000010593	14632
CRM_15	7.45792E+48	3833	HSBA_8	0.999999664	13998
CRM_16	3.63078E+16	2433	HSBA_9	1.000000385	22168
CRM_17	413416987.3	2804	HSBA_10	0.999996953	10473
CRM_18	18.60703382	2769	HSBA_11	1.00000523	10647
LI_7	21.21242645	1941	HSBA_14	1.000017024	10781
LI_8	79.1050169	2431	HSBA_15	1.000000479	13478
LI_9	19.34405777	3157	HSBA_16	0.999999899	12246
LI_10	22.50155045	3321	HSBA_17	1.000000423	12155
LI_11	1.56433E+13	3287	HSBA_18	1.000000937	15820
LI_14	2.1409E+18	2567	RMG_7	1.000010593	4719
LI_15	59784486220	2200	RMG_8	0.999999664	6179
LI_16	1.10331E+55	2219	RMG_9	1.000000385	6317
LI_17	36851577933	3131	RMG_10	0.999996953	5329
LI_18	13.14660649	1351	RMG_11	1.00000523	4047
MMT_7	352.2550892	397	RMG_14	1.000017024	4153
MMT_8	2.927589302	462	RMG_15	1.000000479	5418
MMT_9	11.48763966	913	RMG_16	0.999999899	5174
MMT_10	5.9426E+209	533	RMG_17	1.000000423	6077

MMT_11	8.9095E+15	410	RMG_18	1.000000937	4994
MMT_14	49.28693252	761	SVT_7	1.000000588	4145
MMT_15	21.18898254	778	SVT_8	1.000000042	7358
MMT_16	5.178599144	710	SVT_9	1.000000029	5724
MMT_17	4.73253E+38	1167	SVT_10	1.000000982	7006
MMT_18	15.28643631	517	SVT_11	1.000000951	5241
SVT_14	1.000000774	4564			
SVT_15	1.000000836	5033			
SVT_16	1.000000052	7595			
SVT_17	1.000000501	6614			
SVT_18	1.000000022	8330			
TLW_7	1.000018699	8581			
TLW_8	1.000079081	10009			
TLW_9	1.000031064	6012			
TLW_10	1.000115979	6677			
TLW_11	1.000192133	5060			
TLW_14	0.999984555	9083			
TLW_15	1.000071252	7003			
TLW_16	1.000009491	14029			
TLW_17	1.000253291	8742			
TLW_18	1.000157936	4698			

In Table 2.3, there is extreme variance in the values of the signal-to-noise ratios in the sample. The highest is 4.8843E+231. This, at the prima facie level, appears to be associated with shares listed on NYSE and can be understood in terms of the large trading volumes; however, the danger of the Kalman filter is that, if the updates are not normally distributed, then the estimate is not a best estimator. The complexity is that financial data is not distributed over an even series of intervals, and it is possible for the price feed to contain 'jumps' that can be captured in terms of the trend (Haykin, 1996). The following section of this chapter examines the nature of idiosyncratic effects.

However, it is possible to achieve an overstated purity if the signal series were to possess a moving-average component. If this component is not visible in the correlogram when the autoregressive order is discovered, a type one error (false positive) may result. The issue here is the robustness of methods in terms of their ability to effectively distinguish an autoregressive $AR(p)$ process from an autoregressive moving-average $ARMA(p,q)$ process. It is still important that the issue predominantly applies to stocks listed on NYSE, and the idiosyncratic effects are significant, as established in Model B in section 2.4.4. This is evidence that methods that are valid in one marketplace may not be effective in another.

2.5.4 Data analysis

2.5.4.1 Specification 1

It is important to examine the relationship between the number of bull and bear turns in prices and the total number of orders (regardless of volume) submitted in the trading day. It should be possible to comment on the nature of market efficiency, as if it is the case that bull and bear turns have a meaningful relationship to the market activity. This would suggest that patterns in prices do represent information; ergo, insignificant explanatory variables in this model are an indicator of an efficient market.

The estimated equation is:

Equation 8

$$\boldsymbol{Obs_{it} = \alpha + \beta_1\, Bull_{it} + \beta_2\, Bear_{it} + \varepsilon_{it}}$$

where *Bull* denotes the number of upward price turns and *Bear* denotes the number of downward price turns for asset i on day t.

This is an ordinary-least-squares panel model with random effects. The random effects specification is theoretically justified by the existence of many 'black-box' factors explaining day-to-day differences in trading activity. By applying the random effects, the relationship may become overstated (the Hausman test supports this position). The estimation output is shown in Table 2.4 below.

Table 2.4 Specification 1 estimation output

Variable	**1** **OLS-RE** **coefficient**
Intercept	6896.98
BULL	46.79*
	(28.61)
BEAR	-42.67
	(28.87)
N	1220299
Sample	7Jan–18Jan
Cross-sections	15
Fixed effect	No
R-squared	0.063

*NB: * denotes significance at the 10% level only*

Standard errors in parentheses

Before interpreting the coefficients, it is important to note that the explanatory variables are statistically insignificant at both 5% and 10% significance levels. The R-squared statistic for the model is 0.063, suggesting a very poor specification.

Such an insignificant model lends support to the concept of efficient markets, as it is clear that the existence of reversal shapes (commonly used in short-term technical analysis) generally does not affect the number of orders submitted to the exchange for an equity.

However, when interpreting the parameter values, the sign of the coefficients is of interest, although caution is necessary given the poor fit and power of the model. The signs alone would suggest that bull turns lead to greater activity and bear turns dull interest in the market. Existing literature does not seem to explain this well. Given the limited significance level attached to the finding replication of the finding using data from other time periods and assets would be needed to reach a defensible conclusion in this matter.

2.5.4.2 Specification 2

Some amount of the noise on exchanges may be specific to the exchange rather than to the individual share. When presented graphically, it is possible to 'eyeball' a common component and individual variances. This individual variance can be illustrated by plotting the values calculated for the HKEX listed shares, as shown in Figure 2.5.

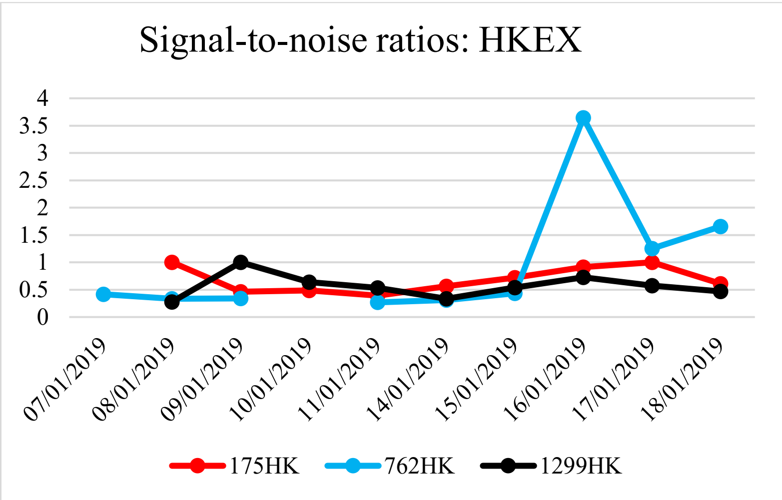


Figure 2.5 Signal-to-noise ratios for HKEX

In order to capture what effects may exist at the exchange level, the sample has been provided with indicator variables to allow for the estimation of a logistic regression (the logit model). For the four exchanges, only three indicator variables are needed, and an interpretation of their coefficients is relative to the excluded variable. In this instance, the base exchange is LSE.

Equation 9

$$\text{Log(Ratio)}_t = \alpha + \beta_1 \text{HongKong}_t + \beta_2 \text{Paris}_t + \beta_3 \text{NewYork}_t + \varepsilon_t$$

In addition, the Huber-White heteroscedasticity-consistent standard errors have been used to correct heteroscedasticity in the model that arise due to the non-independence of the error terms between the exchanges captured in the explanatory variables. This finding indicates that there is a level of global information that possibly captures large macroeconomic factors. However, by controlling for location, it is possible to undertake accurate hypothesis tests to show the significance of the exchange-specific variances.

The estimated output is shown below in Table 2.5.

Table 2.5 Estimation output of Specification 2

Variable	**2** **OLS-RE** **Coefficient**
Intercept	0.0000209
Hong Kong	-0.516 ***
	(0.114)
Paris	45.98 ***
	(17.13)
New York	113.45 ***
	(23.08)
N	1220299
Sample	7Jan–18Jan
Cross-sections	15
Fixed effect	No
R-squared	0.272

*NB * , **, *** Denotes significance at the 10%, 5%, and 1% levels respectively.*

Standard errors in parentheses

The indicator variables are statistically significant at a confidence level of 99%. In this case, the significance of the explanatory variables confirms that the exchange on which an asset is traded is a significant factor in the sonority of the signal within the price feed. Thus, any process that is effective on one exchange may not be effective when used on another exchange. This conclusion highlights the need to understand how exchange rules and cultures can affect the way in which strategies are formed and how information is traded into prices.

2.5.4.3 Specification 3

It is necessary to consider the proposition in the study of Easley et al. (1996) that trading puts private information into prices and that the price feeds should be less noisy when the trading volume is higher, as the incorporation of new private information is more effective and gradual (the series should contain fewer jumps).

Given the calculation of signal-to-noise ratios, this becomes a testable hypothesis in a panel model. Due to the results obtained in Specification 2, it is necessary to adopt a panel estimation with fixed effects in order to ensure that asset-level effects are observed and that exchange-level effects are isolated.

Equation 10

$$\Delta \mathrm{Log(Ratio)}_{it} = \alpha_i + \beta_1 \Delta \mathrm{Log(Volume)}_{it} + \varepsilon_{it}$$

	3
	OLS-FE
Variable	**Coefficient**
Intercept	138.31
Dlog(Volume)	-6.98 **
	(0.045)
N	1220299
Sample	7Jan–18Jan
Cross-sections	15
Fixed effect	Yes
R-squared	0.031

, denote significance at the 10% and 5% levels, respectively.*

Standard errors in parentheses

Figure 2.6 Panel estimation output – Specification 3

The specification yields a low explanatory effect, as indicated by the coefficient of determination (R-squared). The explanatory variable has a negative sign; ergo, when trading volume increases for an asset, the signal-to-noise ratio should decrease, indicating that the proportion of noise increases as the volume increases. It is noteworthy that a lower ratio denotes a greater amount of noise per unit signal.

Higher signal-to-noise ratio values may be indicative of a greater number of uniformed traders in the market, possibly augmenting the argument made in Easley et al. (1996) and supporting models such as the one found in Black (1986), if it is inferred that volume increases result in an increase in the proportion of uninformed traders. As mentioned in Easley et al. (1996), stocks with greater volumes receive greater analytical coverage, which could explain the increased attention from uninformed traders. However, current research does not support these findings. This conclusion indicates that more research is needed in this regard.

2.5.5 The stability of an asymmetric game

It is possible to consider the impact of the estimates of the signal-to-noise ratios on trading strategies. The strategy that assumes that information is correct must be evaluated in terms of its stability, regardless of the factors affecting decision-making competence. To achieve this goal, it must be proven that this strategy yields (within a game theory environment) a trembling-hand stable Nash equilibrium.

This concept is a refinement of the Nash equilibrium, where it is assumed that a dominated strategy is not played in successive iterations of the game; as a result, the strategy would be evolutionarily stable. The 'stability implies Nash' proposition is employed here, as it is in Samuelson and Zhang (1992 [Theorem Six]).

Here is the theorem as stated:

Assume that (x*, y*) is not a Nash equilibrium. For a player, a strategy is set, such that:

Equation 11

$$x_i^* > 0$$

$$\pi_1(i, y^*) < \pi_1(k, y^*)$$

Values of x* and y* exist, such that:

Equation 12

$$\frac{\dot{x}_i}{x_i} < \frac{\dot{x}_k}{x_k} - \delta$$

Here, $\delta > 0$ denotes the probabilistically weighted alternate.

Now, developing a short game to examine the effects of δ, the trembling hand in this case is set as an alternate state of the world (Samuelson and Zhang, 1992).

In this game, two subgames exist, and a state from nature randomises between signal and noise with the above probabilities. It should be the task of a trader to determine if their feed is noise or signal; however, at this stage it is not necessary to be able to tell the difference.

To prove this position, a simplified game must be constructed:

Players

Consider two players, who play asymmetrically. Player 1 is a HF trader who receives a first mover's advantage of (+1) when he acts. Player 2 is an EE trader who has lower payoffs, as they only gain the first mover's advantage when Player 1 chooses to pass, while Player 2 chooses to act.

Strategy set

Both players share the same strategy set: They can react, eliminate, or pass. Reacting incorporates information into price and, in terms of payoff, involves revising holdings. Elimination involves detecting an arbitrage opportunity and acting to remove the noise. A player can also choose to remain idle and pass.

Payoffs

The correct action returns three utils, and the other action returns two utils. Inaction returns no payoff. Note the details of thee first mover's advantage, as outlined above.

Structure of the game

Nature dictates noise or signal. The players act simultaneously.

Based on the Kalman filter results, it is possible to establish the proportion of the price feed that is statistically noise; thus, it is possible to calculate the probability of a movement in price being signal or statistical noise. This data is available in a table in Appendix 2.3.

In order to test the trembling-hand stability of the Nash equilibria in Specification 3, it is necessary to assign numerical payoffs in order to make the test. Thus, example values of the parameters must be used. The model is as outlined in section 2.3.3 and requires given values for *s, x*, and *z* to make a test using the theory outlined in this present section.

Take the following as exogenous:[10]

- The spread to be closed (x) is \$0.01.
- The proportion of the spread closed by the first trader (s) is 0.6.
- The cost per trade (z) is \$0.001.

These values yield the following:

Equation 13

$$\pi = (x \cdot s) - z = \$\mathbf{0.0050}$$

Equation 14

$$\pi = (x \cdot (1 - s)) - z = \$\mathbf{0.0030}$$

And the reduced payoff for sub-penny jumping is:

Equation 15

$$\pi = ((x - 0.0001) \cdot s) - (z) = \$\mathbf{0.00494}$$

Here, it is only necessary to calculate payoffs that are within the candidate equilibria.

The calculations of the payoffs multiplied by the probabilities are available in a table for all sample members in Appendix 2.3.

Examining equilibria (HR, LR)

First, Player 1's choice is between reaction and non-reaction. The non-reaction option causes Player 1 to receive a payoff of zero, and the positive expected profit will lead to acceptance of the reaction strategy. Player 2 has the option to reject the first payoff in order to achieve the second payoff. This is not viable unless the value of *s* is less than 0.5. In this example, where *s* equals 0.6, Player 2 is not willing to defect. As a

[10] These values are based upon examples provided by the EUREX trading venue. An assumption is made here that traders satisfy appropriate position checks before submitting orders.

result, (*HR, LR*) is a trembling-hand perfect Nash equilibrium in the case of all the observed probabilities of signal in the sample.

It is noteworthy that the factor that would be most destabilising is the volume (within term *s*) is the greatest. As this is a simplified version of reality, the conclusion is that a rational trader would attempt to close the greatest proportion of the spread as possible, and they increase their profits by doing so. It is not necessary to close more than half of the spread by volume as is necessary in the dual-trader environment.

Examining equilibria (LR, HS)

Player 1 is making a choice between the second payoff and a payoff of zero. As long as Equation 6 produces a positive value (in all the sample cases), Player 1 will not defect and will always play a reaction strategy.

While the value of *s* is less than 0.5, Player 2 will not defect from the sub-penny jumping strategy to the limit-order strategy. The comments above regarding the value of *s* also apply in this situation. (*LR, HS*) is thus a trembling-hand perfect Nash equilibrium in the case of all the observed probabilities of signal in this sample.

As a result, the probability of trading not being based on information does not affect the stability of the models' solutions, and it is a viable strategy for trade, regardless of the legitimacy of the signal sonority when using spread-capture strategies.

It is important to note that this model does not reflect the cases of long-hold strategies – it merely proves that, when scalping over very short periods, a risk-neutral investor need not make great effort to gain information to ensure a trade is made on signal rather than statistical noise. There is no need to make a distinction.

2.6 Conclusion

This chapter set out to evaluate the trading environment and update the microstructure models to allow for research which considers multiple trader types. This chapter has clarified the variances in trading activity between assets and exchanges and has exposed the logic underpinning the need to participate in the 'arms race' to reduce latency and use non-standard order types.

Trading environments characterised by multiple trader types are also characterised by variations in the latencies of traders, who are required to non-cooperatively interact with each other. This variation is marked by latencies that differ, which allow, for faster algorithmic traders, the possibility of using a greater number of trading strategies. Additional issues are created when the role of a designated market maker is taken by a HF trader who withdraws liquidity before slower traders are able to access it (Stenfors and Susai, 2019).

Using a partial asymmetry model, this chapter has depicted a simplified view of a market of only two traders who are reacting to the same information. The importance of latency in determining the strategy that a player is best incentivised to adopt has also been shown. When a fast trader does not possess the first mover's advantage, they can sub-penny jump to achieve it; however, when a trader does not have a latency advantage over a trader ahead of them, then the best response is to use a limit-order strategy, as doing so imposes the lowest trading cost. The Nash equilibrium outcomes are trembling-hand stable in the presence of noise within the price feeds, as detected by the Kalman filter.

The estimates of noise in price feeds (generated using the Kalman filter) show that the trading venue is a significant factor in explaining noise in prices, and some variances in assets are attributable to the nature of each exchange. In addition, it is possible to detect small chartist shapes within price feeds, which have been shown to have no effect in accordance with efficient market principles (Roberts, 1959). However, these shapes are worthy of further investigation, as they may be of use to a very low-latency trader with a short time horizon.

Following this chapter, it is necessary to better understand how liquidity flows through markets and what reaction times are necessary for traders to react to information. It would also be desirable to illustrate noisy amplification with data. Such analysis could use HF data. These issues are addressed in Chapter 3.

This chapter has not directly considered fairness and the regulatory environment, which form the basis of a survey and discussion in Chapter 4.

Appendix 2.1. Kalman filter specification

The specification is as outlined in Chapter 2. Below is the EViews syntax used to effect the specification, as derived from examples shown in the studies of Bossche (2011) and Rummel (2015). In this case, the daily price data is loaded into EViews and is called by the syntax in line six.

param c(1) -10 c(2) -10

@ename e1

@ename v1

@evar var(e1) = exp(c(1))

@evar var(v1) = exp(c(2))

@signal **Variable_Name** = trend + e1

@state trend = trend(-1) + v1

Figure 6.7 An example of the syntax for the Kalman filter estimation in the EViews state space specification window

Appendix 2.2. Model C solution

Table 6.6 Model in the normal form. The values in red denote the best response, and the cells shaded in pink denote NE by mutual best response.

N: Signal is **HIGH**

1 / 2	L, R	L, F	L, S	L, P	H, R	H, F	H, S	H, P
L, R	1, 2	2, 3	2, 4	1, 0	1, 2	2, 3	2, 4	1, 0
L, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0
H, R	2, 1	3, 2	4, 2	1, 0	1, 2	2, 3	2, 4	1, 0
H, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0

N: Signal is **LOW**

1 / 2	L, R	L, F	L, S	L, P	H, R	H, F	H, S	H, P
L, R	1, 2	2, 3	2, 4	1, 0	1, 2	2, 3	2, 4	1, 0
L, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0
H, R	2, 1	3, 2	4, 2	1, 0	1, 2	2, 3	2, 4	1, 0
H, P	0, 1	0, 3	0, 4	0, 0	0, 1	0, 3	0, 4	0, 0

Strategies and payoffs are as defined in Chapter 2.

Appendix 2.3. Trembling-hand stability proofs

Table 6.7 Probability-weighted payoffs from Model C

Code_date	Signal-to-noise ratio	Probability of signal	Pi 2	Pi 1	Pi 4
175_8	1.00072729	0.500181756	0.001500545	0.002501	0.002470898
175_9	0.46506877	0.317438184	0.000952315	0.001587	0.001568145
175_10	0.484136256	0.326207418	0.000978622	0.001631	0.001611465
175_11	0.392047604	0.281633762	0.000844901	0.001408	0.001391271
175_14	0.560829544	0.359315049	0.001077945	0.001797	0.001775016
175_15	0.721658407	0.419164687	0.001257494	0.002096	0.002070674
175_16	0.914329003	0.477623753	0.001432871	0.002388	0.002359461
175_17	1.000349732	0.500087418	0.001500262	0.0025	0.002470432
175_18	0.610425097	0.379045942	0.001137138	0.001895	0.001872487
762_7	0.416571161	0.294070056	0.00088221	0.00147	0.001452706
762_8	0.334160229	0.250464841	0.000751395	0.001252	0.001237296
762_9	0.341219385	0.254409822	0.000763229	0.001272	0.001256785
762_11	0.268606302	0.211733381	0.0006352	0.001059	0.001045963
762_14	0.311651542	0.237602391	0.000712807	0.001188	0.001173756
762_15	0.433900629	0.302601603	0.000907805	0.001513	0.001494852
762_16	3.645475124	0.784736766	0.00235421	0.003924	0.0038766
762_17	1.253767541	0.556298517	0.001668896	0.002781	0.002748115
762_18	1.656115745	0.623510383	0.001870531	0.003118	0.003080141
1299_8	0.279228189	0.21827864	0.000654836	0.001091	0.001078296
1299_9	1.000311441	0.500077848	0.001500234	0.0025	0.002470385
1299_10	0.639724587	0.390141486	0.001170424	0.001951	0.001927299
1299_11	0.532695208	0.347554559	0.001042664	0.001738	0.00171692
1299_14	0.336042935	0.25152106	0.000754563	0.001258	0.001242514
1299_15	0.539293203	0.350351188	0.001051054	0.001752	0.001730735
1299_16	0.727404072	0.421096652	0.00126329	0.002105	0.002080217
1299_17	0.574597357	0.364917008	0.001094751	0.001825	0.00180269

1299_18	0.47087493	0.320132542	0.000960398	0.001601	0.001581455
AAPL_7	3.60697E+26	1	0.003	0.005	0.00494
AAPL_8	5.11619E+20	1	0.003	0.005	0.00494
AAPL_9	6.7257E+118	1	0.003	0.005	0.00494
AAPL_10	1.44144E+76	1	0.003	0.005	0.00494
AAPL_11	5.64379E+85	1	0.003	0.005	0.00494
AAPL_15	65765157874	1	0.003	0.005	0.00494
AAPL_16	4.8843E+231	1	0.003	0.005	0.00494
AAPL_17	1.06184E+64	1	0.003	0.005	0.00494
CRM_7	1.61452E+18	1	0.003	0.005	0.00494
CRM_8	1.001521973	0.500380204	0.001501141	0.002502	0.002471878
CRM_9	8.1967E+17	1	0.003	0.005	0.00494
CRM_10	1.05111E+78	1	0.003	0.005	0.00494
CRM_11	4.45062E+13	1	0.003	0.005	0.00494
CRM_14	482.965795	0.997933738	0.002993801	0.00499	0.004929793
CRM_15	7.45792E+48	1	0.003	0.005	0.00494
CRM_16	3.63078E+16	1	0.003	0.005	0.00494
CRM_17	413416987.3	0.999999998	0.003	0.005	0.00494
CRM_18	18.60703382	0.948997895	0.002846994	0.004745	0.00468805
LI_7	21.21242645	0.954980155	0.00286494	0.004775	0.004717602
LI_8	79.1050169	0.987516387	0.002962549	0.004938	0.004878331
LI_9	19.34405777	0.950845598	0.002852537	0.004754	0.004697177
LI_10	22.50155045	0.957449616	0.002872349	0.004787	0.004729801
LI_11	1.56433E+13	1	0.003	0.005	0.00494
LI_14	2.1409E+18	1	0.003	0.005	0.00494
LI_15	59784486220	1	0.003	0.005	0.00494
LI_16	1.10331E+55	1	0.003	0.005	0.00494
LI_17	36851577933	1	0.003	0.005	0.00494
LI_18	13.14660649	0.929311669	0.002787935	0.004647	0.0045908
MMT_7	352.2550892	0.997169184	0.002991508	0.004986	0.004926016
MMT_8	2.927589302	0.745390894	0.002236173	0.003727	0.003682231

MMT_9	11.48763966	0.919920816	0.002759762	0.0046	0.004544409
MMT_10	5.9426E+209	1	0.003	0.005	0.00494
MMT_11	8.9095E+15	1	0.003	0.005	0.00494
MMT_14	49.28693252	0.980114118	0.002940342	0.004901	0.004841764
MMT_15	21.18898254	0.954932589	0.002864798	0.004775	0.004717367
MMT_16	5.178599144	0.838151015	0.002514453	0.004191	0.004140466
MMT_17	4.73253E+38	1	0.003	0.005	0.00494
MMT_18	15.28643631	0.938599213	0.002815798	0.004693	0.00463668
NLSN_7	6.84356E+11	1	0.003	0.005	0.00494
NLSN_8	1.21141E+19	1	0.003	0.005	0.00494
NLSN_9	7.6305E+160	1	0.003	0.005	0.00494
NLSN_10	1.83576E+13	1	0.003	0.005	0.00494
NLSN_11	1.87535E+87	1	0.003	0.005	0.00494
NLSN_14	1.60558E+41	1	0.003	0.005	0.00494
NLSN_15	1.0472E+17	1	0.003	0.005	0.00494
NLSN_16	2.4273E+63	1	0.003	0.005	0.00494
NLSN_17	5.1416E+32	1	0.003	0.005	0.00494
NLSN_18	6.49901E+40	1	0.003	0.005	0.00494
VCT_7	2.9304E+17	1	0.003	0.005	0.00494
VCT_8	6.764E+147	1	0.003	0.005	0.00494
VCT_9	6.2788703	0.862616044	0.002587848	0.004313	0.004261323
VCT_10	30.86948306	0.96862202	0.002905866	0.004843	0.004784993
VCT_11	3.775405691	0.790593708	0.002371781	0.003953	0.003905533
VCT_14	2.15564E+11	1	0.003	0.005	0.00494
VCT_15	6.614109234	0.868664873	0.002605995	0.004343	0.004291204
VCT_16	96.14018144	0.989705599	0.002969117	0.004949	0.004889146
VCT_17	3.32892E+13	1	0.003	0.005	0.00494
VCT_18	1.17651E+15	1	0.003	0.005	0.00494
YUM_7	1.25292E+14	1	0.003	0.005	0.00494
YUM_8	6.78574E+30	1	0.003	0.005	0.00494
YUM_9	1.8331E+39	1	0.003	0.005	0.00494

YUM_10	7.86282E+12	1	0.003	0.005	0.00494
YUM_11	1.000702522	0.500175569	0.001500527	0.002501	0.002470867
YUM_14	52.88284482	0.981441217	0.002944324	0.004907	0.00484832
YUM_15	1074.103165	0.999069857	0.00299721	0.004995	0.004935405
YUM_16	7.04102E+91	1	0.003	0.005	0.00494
YUM_17	38.88522004	0.974928056	0.002924784	0.004875	0.004816145
YUM_18	1.89573E+62	1	0.003	0.005	0.00494
BAE_7	1.000017299	0.500004325	0.001500013	0.0025	0.002470021
BAE_8	1.00000014	0.500000035	0.0015	0.0025	0.00247
BAE_9	1.000001082	0.50000027	0.001500001	0.0025	0.002470001
BAE_10	0.999997536	0.499999384	0.001499998	0.0025	0.002469997
BAE_11	1.000015428	0.500003857	0.001500012	0.0025	0.002470019
BAE_14	1.000010568	0.500002642	0.001500008	0.0025	0.002470013
BAE_15	0.999995116	0.499998779	0.001499996	0.0025	0.002469994
BAE_16	0.999999999	0.5	0.0015	0.0025	0.00247
BAE_17	1.000000639	0.50000016	0.0015	0.0025	0.002470001
BAE_18	1.000000552	0.500000138	0.0015	0.0025	0.002470001
HSBA_7	1.000010593	0.500002648	0.001500008	0.0025	0.002470013
HSBA_8	0.999999664	0.499999916	0.0015	0.0025	0.00247
HSBA_9	1.000000385	0.500000096	0.0015	0.0025	0.00247
HSBA_10	0.999996953	0.499999238	0.001499998	0.0025	0.002469996
HSBA_11	1.00000523	0.500001307	0.001500004	0.0025	0.002470006
HSBA_14	1.000017024	0.500004256	0.001500013	0.0025	0.002470021
HSBA_15	1.000000479	0.50000012	0.0015	0.0025	0.002470001
HSBA_16	0.999999899	0.499999975	0.0015	0.0025	0.00247
HSBA_17	1.000000423	0.500000106	0.0015	0.0025	0.002470001
HSBA_18	1.000000937	0.500000234	0.001500001	0.0025	0.002470001
RMG_7	1.000010593	0.500002648	0.001500008	0.0025	0.002470013
RMG_8	0.999999664	0.499999916	0.0015	0.0025	0.00247
RMG_9	1.000000385	0.500000096	0.0015	0.0025	0.00247
RMG_10	0.999996953	0.499999238	0.001499998	0.0025	0.002469996

RMG_11	1.00000523	0.500001307	0.001500004	0.0025	0.002470006
RMG_14	1.000017024	0.500004256	0.001500013	0.0025	0.002470021
RMG_15	1.000000479	0.50000012	0.0015	0.0025	0.002470001
RMG_16	0.999999899	0.499999975	0.0015	0.0025	0.00247
RMG_17	1.000000423	0.500000106	0.0015	0.0025	0.002470001
RMG_18	1.000000937	0.500000234	0.001500001	0.0025	0.002470001
SVT_7	1.000000588	0.500000147	0.0015	0.0025	0.002470001
SVT_8	1.000000042	0.50000001	0.0015	0.0025	0.00247
SVT_9	1.000000029	0.500000007	0.0015	0.0025	0.00247
SVT_10	1.000000982	0.500000246	0.001500001	0.0025	0.002470001
SVT_11	1.000000951	0.500000238	0.001500001	0.0025	0.002470001
SVT_14	1.000000774	0.500000193	0.001500001	0.0025	0.002470001
SVT_15	1.000000836	0.500000209	0.001500001	0.0025	0.002470001
SVT_16	1.000000052	0.500000013	0.0015	0.0025	0.00247
SVT_17	1.000000501	0.500000125	0.0015	0.0025	0.002470001
SVT_18	1.000000022	0.500000005	0.0015	0.0025	0.00247
TLW_7	1.000018699	0.500004675	0.001500014	0.0025	0.002470023
TLW_8	1.000079081	0.50001977	0.001500059	0.0025	0.002470098
TLW_9	1.000031064	0.500007766	0.001500023	0.0025	0.002470038
TLW_10	1.000115979	0.500028993	0.001500087	0.0025	0.002470143
TLW_11	1.000192133	0.500048029	0.001500144	0.0025	0.002470237
TLW_14	0.999984555	0.499996139	0.001499988	0.0025	0.002469981
TLW_15	1.000071252	0.500017812	0.001500053	0.0025	0.002470088
TLW_16	1.000009491	0.500002373	0.001500007	0.0025	0.002470012
TLW_17	1.000253291	0.500063315	0.00150019	0.0025	0.002470313
TLW_18	1.000157936	0.500039481	0.001500118	0.0025	0.002470195

3 Characteristics of high-frequency trading

3.1 Abstract

High-frequency (HF) traders can trade using strategies that rely on their low-system latency, giving rise to actions that may lead to allegations that HF traders act in ways that are improper. Regulators and researchers who are independent of exchanges require techniques to work with the limited available data formats to examine events of interest in order to evaluate how traders are contributing to the price formation process. Previous research has provided techniques to flag potential HFT activity; however, little is known about identifying individual trading entities' contribution to the price formation process. This research proposes a technique that uses the ARCH qualities of financial time series data to identify the contribution individual traders make to price formation rather than to identify an HFT. The approach used herein is practical, as the definition of HFT is relative, and it is hard to justify existing methods that simply seek to flag HFT activity. Rather, by identifying relevant traders and examining the effects they have on price formation processes, it is possible to deconstruct price volatility moments and examine the effects of individuals on price stability. Results based on a HF ForEx dataset demonstrate that it is possible to single out which traders are significantly contributing to price formation and which traders resist price movement. Along with existing HF detection methods, this proposal forms a robust toolkit for analysing ex-post data.

3.2 Introduction

This chapter seeks to explore how HFT activity may be detected and how market data may be deconstructed with the goal of learning how individual traders contribute to the market price formation process. It begins by surveying what is known about the inner workings or the decision-flow process of a typical HFT system. This survey is particularly interesting, as it goes some way towards explaining how these traders may not use 'all available information' to make a decision, and this limited use of information could also account for some if the risk adverse behaviour observed during market volatility events.

Thus, it is important to consider the methods of detecting HFT activity that are currently in use in the context of the limitations of various datasets. This contribution makes no unique attempt to identify HFT activity; rather, a methodology is proposed to identify the contribution of each individual trader to the market price formation process. This methodological approach takes the form of an ARCH model-based approach that incorporates a working price level for every trading identity in the dataset. In turn, this model allows for individual trading entities to be added as an explanatory variable in the auxiliary ARCH regression equation. This produces a viable tool for retrospective analysis, which is useful for analysing market volatility events or the price formation process in general.

HFT has become a product of continuous development in ICT since the early 1980s (Chlistalla, 2011; Carbó-Valverde, 2017). This progress is likely to intensify in future decades as ICT's ability to process information and mimic floor-trading strategies that have been used for centuries is increasing, along with an increased focus on the latency aspect. This fintech boom has the potential to disrupt current market structures and create new challenges in data governance, adequacy of regulation, and legal accountability (Santiago Carbó-Valverde, 2017). These challenges are likely to be compounded by the effects of increasing competition between trading venues (Chlistalla, 2011). Much attention has been given to HFT following the May 2010 flash crash; while this was not directly caused by HFT activity, questions were raised concerning how beneficial HFT is for markets and how other (slower) traders can interact with HFT. Advances in ICT and liberalised exchange rules have created the opportunity to access markets without the need for traders to be physically present at the venue, which in turn has created a 'digital arms race'[11] (Chlistalla, 2011). In addition, since the 1980s, markets have slowly become more fragmented, especially in the case of equity, as new small venues have emerged and eroded the once legalised monopoly of key exchanges. This phenomenon has resulted in exchanges becoming more competitive in fees, which is largely the reason for the use of the 'maker-taker' pricing/rebate structure – designed to incentivise liquidity. This innovation may not be the last in this area. The maker-taker pricing/rebate structure gives rise to one of the most common areas of activity – a form of arbitrage. Arbitrage is a strategy that relies on simultaneously buying and selling an asset in two different places. Passive

[11] This source seems to omit the greatest boon for remote access and algorithmic generation: the removal of rules that required orders to be entered via the venue's keyboard (many newly digitised exchanges had such rules).

rebate arbitrage is very common in the maker-taker pricing/rebate structure. Furthermore, many HFT systems are known to and have even been designed to undertake latency arbitrage strategies.

An important concept is that the arbitrager who is not the first to reduce a spread to zero or toward zero does not earn a profit. A good way to express this idea is given by Angel (2014, p. 272): 'an arbitrager who comes in second in the race for a profitable trade still loses, whether by one minute, one second, or one nanosecond'. This time pressure is also an incentive for traders to take part in the technological 'arms race', as described by Chlistalla (2011); however, it must be noted that at some point in the future, this latency pressure will be bounded by the fact that nothing is capable of moving faster than the speed of light in a vacuum. Furthermore, in the future, as the latency of markets continues to decrease, there will be another issue: The speed at which information moves between the various members of the exchange (who need not all be in one location) will cause different traders to see different information at one given time (Angel, 2014). Angel (2014) suggests that the various states of the world and asymmetric information creates a Schrödinger's cat-type problem. In the case of Schrödinger's stock market, a trader might not realise the actual state of a market until after an order has been submitted. This situation restricts the perceived fairness of markets and may also present a challenge in regulating markets.

Concerning latency arbitrage, lower latency equates to faster speed, which is brought about by advancing technology (itself spurred on by the 'technological arms race'). In order to undertake latency arbitrage, a firm must process data as fast as possible, usually with the aim of predicting order flow or order volume in the next second (or, possibly, in the next few seconds). Then, the objective is to get the order flow into the exchange's matching engine as soon as possible to scalp, close a spread, or access best execution (Arnuk and Saluzzi, 2009; Hagströmer and Nordén, 2013).

Furthermore, speculative liquidity provision is also a known activity, which must be interpreted against the findings of Stenfors and Susai (2019), where it is illustrated that liquidity provision from HF traders is often available for less than 0.2 seconds, and liquidity may reverse within 0.5 seconds. Hagströmer and Nordén (2013) found that values of HFT activity vary – between 63% and 72% of volume originated from HFT in their samples taken in 2011. This activity is likely to have contributed to price discovery, as HF traders are able to react to news events within 2–3 milliseconds, according to an analysis of the NASDAQ exchange in 2007–2008 (Hagströmer and Nordén, 2013, p. 751). The exact strategies and reasons for using HFT vary, but, broadly speaking, activity can be grouped into either market-making or opportunistic strategies, such as arbitrage, order flow anticipation, and momentum ignition (Hagströmer and Nordén, 2013). Order types used to meet the objectives of HFT strategies also vary and have moved beyond the scope of traditional order types like limit orders, market orders, and stop orders. Increasingly, complex compound order types are now valid in many trading venues, as these allow exchanges to compete against each other, as more complex order types offer traders superior economic results, according to Dolgopolov (2014). Examples of compound order types are 'hide not slide' and 'fill or kill' – many others also exist. This increase in the number and complexity of order types may place more pressure upon those serving as brokers who are bound to seek best execution for their clients. This is because traders need to understand all available order types and how their own orders may interact with other orders in the market of different types (Dolgopolov, 2014). Furthermore, the use of complex strategies by HF traders, combined with their low latency, can give slower traders the impression that liquidity can instantly be 'swept away' and that responses to information can be coordinated, as HF traders react to the same signal (Bodek and Shaw, 2012).

It is helpful to examine patent applications, as these are examples of HFT operations detaining the workings of their own HFT systems and placing this information in the public domain. It is unsurprising that such disclosures are rare. A patent filing issued on behalf of the Bank of America gives a very clear insight into the information flow process or decision-making process. This example of the latency-minimising design of HFT systems is seen within a patent application that is available in the public domain (Cohen, 2012), which was granted and awarded US patent number US8130758132. This filing illustrates the use of multiple data processing feeds within feeds that are data inputs. There are distributed queues for filtration by one of many 'processing units' where filtration (likely using AI) sorts information into categories and identifies relevant system users or subscribers. At this stage, consolidated data is filtered a second time in order to meet the needs of individual subscribers. Here, the design of the system appears to process feed data as quickly as possible due to its dualchannel approach; yet, it also uses two sets of filters to sort data into consolidated feeds for individual purposes so that irrelevant information need not be considered by subscribers (Cohen, 2012)[12].

This description is similar to a neural network system of decision layers between input and output data. Neural networks are derived from the human problem-solving process –a clustering of decision nodes that are used in a system of 'feedforward backpropagation' to produce an estimate of target variables based on input variables. The node positions change over time as the system learns. Such systems are able to produce high R-squared values when the input data is consistent (Macchiarulo, 2018). This finding is consistent with the findings of Morris et al. (2009), who believe exchange data feeds are the primary means of communication of changing prices and changing market conditions. However, it should be noted that HF traders likely use more than just the market data feed in order to contextualise information.

In HF environments, the key objective is to interpret the market data feed as quickly as possible along with other sources of information that may depict future volume in the order book. This 'need for speed' is possibly accompanied by market volatility, which is thought to be exacerbated by the characteristics of HFT. Easley et al (2011) analysed the May 2010 flash crash event and were able to illustrate that the rapid price movement was not initially caused by HFT; however, the algorithmic intelligence in use did appear to exhibit risk-adverse behaviour due to a lack of situational awareness, perhaps caused by overreliance on the market data feed as the source of situational awareness. However, in contrast, it was claimed by the Financial Conduct Authority (FCA, 2018) that algorithmic trading firms do appear to cause extreme price movements due to overly aggressive responses when trading in the same direction as a price change.

The FCA supervision summary evidences that increased speed and reduced costs '[amplify] certain risks' (FCA, 2018). These changes present risks that are to be met with regulatory oversight that 'keeps up' with technology as well as a reliance on strong internal governance. A further line of enquiry concerns the effects of HFT on moving price, and how price movement is interpreted by other traders. In such cases,

[12] Figure 6 of the patent application shows a flow diagram that illustrates this.

it could be that price is not reflective of all information. Thus, the efficient market hypothesis may be questioned, at least over short horizons. This idea may be worthy of further research; however, it is beyond the scope of this chapter. Relevant literature suggests that technical analysis may be effective in practice (Chande, 2001). Some evidence suggests that patterns may have some explanatory effect (one example is included in Chapter 2). Lo et al. (2000) examined data over a period of three decades and were able to illustrate that some (but not all) indicators may provide users with 'some practical value' (Lo et al., 2000).

In considering HFT and analysing its behaviours, it is worth noting the nature of 'big data', which is relevant to the characteristics of HF information processing according to Seddon and Currie (2017). The nature of HFT and algorithmic trading is defined by the challenges of big data as a determinant of how these traders action their information set and generate their order flow. Big data typically has three attributes: volume, velocity, and variety. In Seddon and Currie (2017), these are referred to as 'big data', 'fast data', and 'big compute'. It is sometimes the case that AI can aid the performance of HFT by sorting data rather than forcing systems to process the whole data feed at increasingly reduced system latencies. An example of reducing information transmission times (system latency) is the fibre-optic connection between New York and Chicago, linking the Chicago Mercantile Exchange and various exchanges located in New York City (including NYSE). This fibre link significantly reduces the data transmission times compared to the older telephone lines installed in the late 1940s (Laughlin et al., 2014).

Along with generally reduced latencies, a variance in trading motives has been observed by Hagströmer and Nordén (2013). This line of enquiry is extended by Menkhoff and Schmeling (2010) who seek to illustrate 'whose trades convey information'. In order to establish the characteristics of an informed trade and specifically to assess the price impact made by individual traders or groups of traders, a price impact analysis was performed. In order to do so, disaggregated data from Moscow's MICEX exchange was obtained for the Rouble–USD pair for a total of nine days in 2002 (Menkhoff and Schmeling, 2010). Based on extant literature, it was assumed that five factors may be explanatory: trader size, proximity to a financial centre/hub, time of day, prevailing bid-ask spread, and outstanding order book volume. From the dataset, 50,000 price impact observations were detected, and trade size, trader size, and time of day were found to explain price movement (Menkhoff and Schmeling, 2010). The former factor is expected and the latter is supported by Barclay and Hendershott (2003), who illustrated that information incorporation varies across the trading day and the overnight period.

Although the factors influencing price impact are many, varied, and generally known, it is not as easy to forecast price movements using an algorithmic approach. Brown et al. (2013) examined the factors that may be explanatory of price formation when used in an algorithmic approach. These explanatory factors were: population size, number of generations (iterations), mutation/update rate, seed radius, radius delta, revaluation loop count, convergence limit, and weight. All these factors refer to the specification of the algorithm, thus illustrating that it may be possible to detect changes in prices over time if the algorithm is sensitive enough and calibrated to existing information. In addition, based on this algorithm, the factors can self-modify to continue to follow prices. The analysis found that this algorithm can select portfolios of single stocks over a one-week horizon, which performed in excess of the Dow Jones Index (as a benchmark). What is not clear is how far the horizon of predictions is effective.

Predicting price is perhaps a secondary consideration to the ability to predict and/or model volume, given that many algorithmic traders forecast order flow and infer price directionality based on movements in volume. However, Wang (1994) also reminds us that prices should be reflective of aggregate risk and that trading in markets should reflect how information affects the risk associated with holding an asset. As information is an element here, the issue of asymmetric information enters the market. Wang (1994) examines the relationship between volume and prices with the aim of showing how the heterogeneity of investors may affect asset prices. Where asymmetry in information exists, there may be private information against uninformed traders who may require a greater return (or discount in price) in order to justify trading against private information (greater uncertainty equates to greater risk). In short, less informed traders are more cautious when trading against more informed investors because, when a market is in equilibrium, Wang (1994) believes that when an uninformed trader trades, they will display any private information they hold via the bid or ask price. An informed trader will sess the existence of information as a change in price. They, of course, will either accept this as information or could attribute the change in price to a noise component, consistent with the findings of Black (1986).

3.3 The workings of a high-frequency trader

Before looking at the methods that can be used to detect HFT activity and exploring the possibility that machine learning methods may be helpful, it would also be useful to summarise some key information about HFT systems and their operation.

HF traders are low-latency algorithmic trading systems. What exactly represents the speed that is necessary for activity to be considered HFT is not universally defined – it is a relative term (Chlistalla, 2011; Wang and Zheng, 2015). Some reputable sources surveyed in the following sections take the view that if a trading entity can revise its position within 100 milliseconds of an order, it is likely capable of HFT activity. However, speed is not the only characteristic that is relevant – factors such as the order types used can be helpful in detecting HFT, too (Dolgopolov, 2014).

The order types used by HFT systems have been the source of some controversy, as discussed in relevant literature. At the heart of this controversy are the revelations of a former trader named Haim Bodek, who alleged that some order types allow for traders to jump in front of other orders; however, it must be noted that Bodek's contribution cannot be independently verified due to the proprietary nature of the HFT algorithm design and activity. In all probability, he experienced a scenario in which he was trading against traders who were using 'hide-not-slide' order types.

Dolgopolov (2014) explained that order types in common use are often more complex in nature than traditional limit orders (buying at a price or better). For example, a 'hide-not-slide' compound order is designed to act as a limit order that takes on an additional property in the event that the market locks.[13] When this situation occurs, the 'hide not slide' remains in the order book of the exchange and is prioritised when the market unlocks. HF traders face pressure to use these complex order types as they are commonplace among competing firms and failure to use these orders may impinge a brokerage's best-execution obligations (fiduciary duty) to their clients (Dolgopolov, 2014).

At face value, HF traders could be approximated to electronic market-makers; however, in this role, HFT is also known to adopt scalping strategies. These strategies

---

[13] The market is said to lock when the bid-ask spread equals zero. In this case, trading on exchanges halts until the bid and ask separate.

are not inventory neutral; rather, they seek to capture ‘exchange rebates upon electronic exchanges running the maker-taker market model (Bodek, 2013). Scalping is a spread-capture strategy, typically conducted over short time horizons. In HFT, this horizon could be as short as six market ticks, which could occur in less than a second if markets are sufficiently active.

This is how Haim Bodek (2013, p. 18) describes the scalping strategy:

> Its core intent is to, on every round trip, trade to step ahead of supply and demand imbalances evident in market depth and to capture a micro spread by closing on the other side for a tick or to scratch out by closing on the same side, both of which are favourably subsidised by the rebate in the maker-taker market model.

An additional consideration is the possibility of front-running scalpers, who are able to process information faster than the security information system of an exchange. When quotes can be updated faster than the speed of information dissemination, this is a latency arbitrage-based strategy, which likely makes trading more complex for slower traders. As a result of this speed, HFT scalpers aim to gain a favourable queue position. Any strategy must have a sound probability of entering the trade and an equally high probability of being able to exit the trade in order to prevent losses if the spread could not be captured (Manahov, 2016). To a trader, the market depth when scalping is all the orders queuing behind him. An effective HF trader has sufficient speed and skill to get priority in the queue to execution and be the first to buy or sell. In an extreme situation, a scalper trades against those behind them in the queue to execution (Patterson, 2012, p. 52).

Patterson (2012, p. 204) quotes sections of an interview with Dave Cummings, a floor trader turned HF technologist. Cummings created a ‘trading robot’ based on pit-trading principles. He later was employed by Getco. The following quotation gives an insight into the aims of those developing HFT technology.

> “It became about meeting the needs of that specific HFT community,” says a technologist who worked for several top ECNs and exchanges in the 2000s. “The game changed. Firms like Getco and Tradebot wanted to know everything about our system so they could manage their orders

> accordingly. We spent a tremendous amount of money trying to meet their needs – they trained us to be fast. It is all about the functionality I can offer the HFT that they can take advantage of. We're going after guaranteed economics."

Much of the ability to use order types effectively depends on the ability to understand how each individual exchange processes orders and which order types they do or do not allow. A summary of a typical matching flow is shown in Banks (2009).

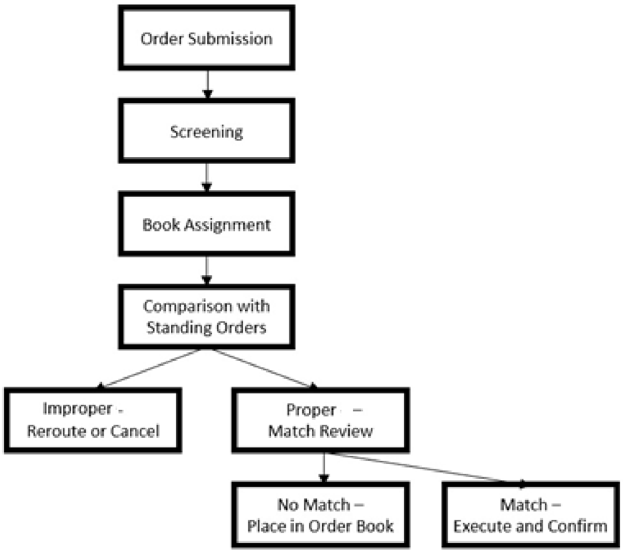


Figure 3.1 Order matching flow (Banks, 2009)

Figure 3.1 shows the screening process that takes place within the exchange. Upon order submission, a screening process takes place in order to prevent a trading entity from submitting multiple identical orders in rapid sequence by error or design. In addition, at this point, checks are made against other counterparties to ensure that two traders are not consistently submitting identical orders. This latter check is a safeguard against collusion and insider dealing. A check is then made against the exchange rules (or standing orders) to ensure that the order type is valid and the initiator has appropriate standing to trade in the particular venue. Thereafter, the order is routed toward an order book, when execution may be possible. The order can then either be crossed immediately (where liquidity exists) or it is added into the order book to await a crossing order (Banks, 2009).

Less is known about the inter-workings of the HF traders who submit their orders to exchanges. Fortunately, it is possible to learn a little from patent filings, in which an applicant is required to describe the nature of what they wish to protect. A filing from Bank of America (Cohen, 2012) details the processing of information by an HF trader by using a series of queues to achieve a low system latency.

HFT patent filings do not appear to be commonplace, as designers face an incentive to protect intellectual property wherever possible; however, those who do seek intellectual property protection are required to detail their innovation and place details in the public domain. This disclosure goes against the typically proprietary nature of HFT design and development. For this reason, patent filings are only detailed enough for the unique element to be recognised but not replicated.

The system's cost/reward behaviour is explained by three equations that outline the costs of running these systems and some of the definitions of the terms used in the system operation.

Equation 16

$$R = \sqrt[n]{\coprod_{1}^{n} Q_i}$$

Here:

R – response time (latency)

Q_i – execution time for query i

The response time or general latency of the system is defined as the geometric mean of the execution time of each query.

Equation 17

$$T = \frac{\sum E_i}{N}$$

This is the average time taken to complete a query.

Here:

T – throughput (volume of queries processed)

E_i – elapsed time to execute one query (query i)

N – number of queries processed in a unit of time

Equation 18

$$C = \frac{R \cdot T}{TC}$$

where C is cost per trade (average cost) and TC is the total cost of the system per unit time – both of these values are monetary amounts (Cohen, 2012). It is possible to conclude from the cost identity that the system faces lower running costs if the latency is lower, and this cost is met by higher values of throughput.

Cohen (2012) explained how information flows through the system and end users connect with a partitioned message board. The aim of the system is to take the exchange data feed and construct an internal message board (pulses of information) upon which decisions are actioned as strategies via applications. It is important to note that the point of this system is to screen data that is sent into applications that create or submit trade ideas.

Figure 3.2 Block diagram illustrating a low-latency data system (Cohen, 2012)

Figure 3.2 illustrates the core element of a single partition for handling data. The system always comprises a feed handler and a message board; however, the application element can vary according to the application of the system. These applications could incorporate machine learning techniques. The feed handler de-encapsulates the data and presents it to the message board. The message board distributes the de-encapsulated data to relevant subscribers using a system of filtering to provide an asynchronous communication mechanism.

This diagram only reflects a small part of a HFT information system, as multiple partitions can run in parallel with the aid of a content-based router. This is shown visually in Figure 3.3 below, which is based on a diagram in the discussed patent application (Cohen, 2012).

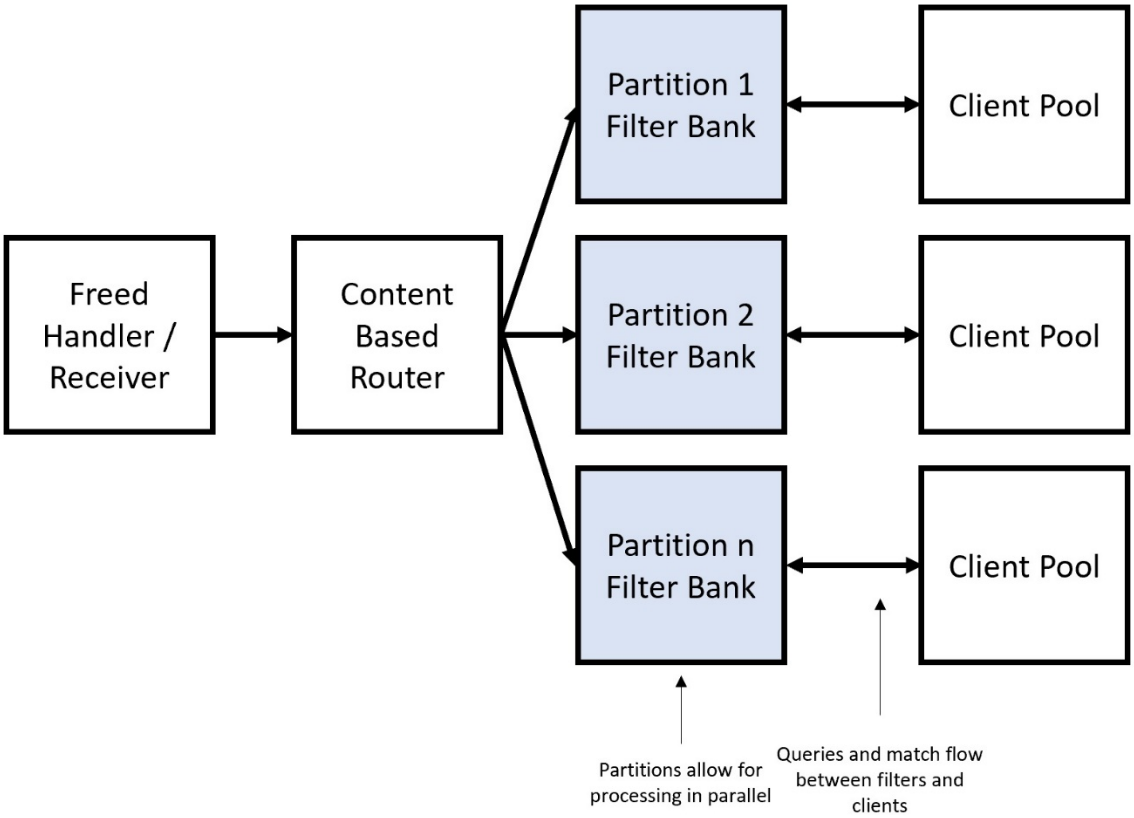


Figure 3.3 Block diagram illustrating an HFT message board with parallel processing capability (Cohen, 2012)

It is within the message board that the data is applied to a particular model. Systems must be able to add and remove capacity as needed in order to produce scalable consistency in their service quality. There are two limitations to this idea: Firstly, the linkage between the feed handler and the message board is complex, and secondly, it is necessary to provide a method to query live data feeds. These are constraints to system latency that are difficult to overcome.

In addition, the message board can be supplemented by additional analytical systems providing value-added analytics, which analyses the content of the message board and actions these back to the message board (Cohen, 2012). It may be concluded that not all information will necessarily become part of the system's situational awareness, which may somewhat explain why these trading systems tend to be deemed risk averse and, at times, appear to overreact to certain extreme events.

What can be learnt from this patent application appears consistent with the findings of Morris et al. (2009), who illustrated exchange data feeds as the primary means of communication of changing prices and changing market conditions. In HF environments, the key objective is to interpret the data feed as quickly as possible, along with other sources of information that may depict future volume in the order book.

An example of what is possible when processing power allows for low latency is given by Wah and Wellman (2013). They demonstrated how a latency arbitrager may operate between two markets. Assuming multiple background traders are trading in two markets and an infinitely fast arbitrager is trading across both markets, the arbitrager with the aid of a best bid and offer feed calculated across both exchanges (NBBO) by an SIP is able to calculate the spread between prices in each market.

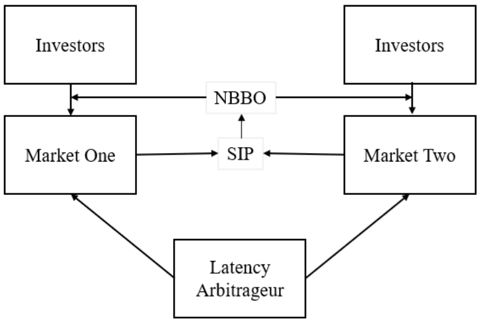


Figure 3.4 Diagram depicting the role of a latency arbitrager in a dual-market scenario (Wah and Wellman, 2013, p. 5). NBBO denotes the national best bid and offer, and SIP denotes the security information processor that calculates the NBBO.

In more practical terms, when latency is sufficiently lower than with other background traders, arbitrage of this nature is possible. It is not necessary to assume an infinitely fast trader as in Wah and Wellman (2013); rather, it is only necessary to assume the arbitrager is of a consistently lower latency than the background traders.

3.4 Measuring and detecting HFT activity

3.4.1 Order-to-trade ratios

Order-to-trade ratios (OTRs) vary according to each trader and are likely to be affected by both trader strategy and trader type. A trader's cancellation of orders before they are executed contributes to an order-to-trade ratio above one.

It is known that HF traders often submit orders and withdraw them before they are executed. Given their processing power and low latency, they are better equipped than any other trader type, if the market is liquid and has a fast rhythm through the order book, to make use of the ability to cancel orders. This appears to be the belief of a selection of trading venues, notably Nasdaq and EUREX, who set restrictions concerning permissible OTR values in order to limit the scope for HFT activity. This form of HFT control is theoretically effective in preventing market abuse using a spoofing strategy or a quote stuffing strategy. Both approaches are intended to flood the exchange matching engine, but the intention is different. Spoofing aims to create the illusion of liquidity which the initiator can trade against. Quote stuffing on the other hand, involves sending spurious orders in repeated bursts in order to slow down the matching process (Wang and Zheng, 2015).

What is lacking in the existing literature and material produced by exchanges is a clear threshold in terms of what OTR values may denote HFT activity on a trader's account. As a trader activities and motivations vary, this may be impossible to define. Exchanges set requirements regarding the OTRs that trading entities achieve while making allowances for the roles of strategic trading and designated market making by defining the ratios differently in each case. The definitions of the ratios appear to be standard, as is the application of the counting principles. However, limits on acceptable values of the OTR vary between exchanges.

Orders that fall into the non-market-making category are summated to provide:

Equation 19

$$ORT = \frac{\sum Orders}{\sum Trades} - 1$$

In the market-making role:

Equation 20

$$ORT = \frac{\sum \frac{Orders}{Quotes}}{\sum Trades} - 1$$

These definitions are given in Nasdaq (2019). The counting methodology provides a framework for determining what is and what is not an order. For example, the submission of a limit order counts as an order submission (+1), and if this limit order is then amended, the count rises to two (1+1). In the instance that a party provides a quote, this has a count value of two: +1 for the bid and +1 for the ask side. It is important to note that combination orders count as value 1, so a hide-not-slide order should only count once.

As of 2 January 2018, the Nasdaq exchange enforced the following maximum permissible values of the OTR (which depends on asset class). As the maxima vary, this enforcement implies that the count is separate for each class and is not combined.

Table 3.1 Maximum order-to-trade ratios on the NASDAQ exchanges (Nasdaq, 2019)

Sub-asset class	**Max. OTR**	**Max. OTR (market-making roll)**
Index futures	150,000	1,500,000
Index options	15,000	15,000,000
Stock options	15,000	10,000,000
Stock futures	150	1,500,000

It is important to note that Nasdaq only applies these to derivative markets and not to their equity listings. Unfortunately, there is no written rationale explaining why this is the case.

Here, the implementation is a little different to Nasdaq, as the system is not based on count; instead, it is based on volume. Therefore, the OTR is:

Equation 21

$$OTR_{vol} = \frac{Order\ Vol}{Trading\ Vol}\ for\ any\ time\ period$$

In the case of a trader whose trading volume is low (below 1,000), the trading volume is subject to a floor value of 1,000. For example, if a trader has a trading volume of 900, the denominator in the formula is 1,000. This is done to minimise the effect of additional orders where trading volume is low, although this is not directly stated (Peters, 2016).

The EUREX group sets threshold values that vary by asset class – like Nasdaq, EUREX only applies the OTR in derivatives markets. In the volume-based system, traders who serve as market makers are permitted to work to increasingly high OTR values as their bid-ask spread widens (EUREX Deutschland, 2021).

EUREX claims that it implemented this system of OTR limits in response to the German HFT Act (Hochfrequenzhandelsgesetz, Germany; Ba Fin, 2013), which overtly seeks to restrict HFT activity. However, EUREX's documentation does not indicate that this limitation on OTRs only affects HF traders (EUREX Deutschland, 2021).

In summary, the OTR value may be an indicator of an HFT; however, it cannot be a diagnostic tool due to the relative nature of the definition of HFT.

3.4.2 High-frequency trading detection methods

3.4.2.1 Dataset

In this section, the following dataset is used to illustrate the HFT flag method and two machine learning techniques.

The dataset is taken from Wednesday, 01 June 2016 and captures the whole trading period. Six separate currency pairs are included: GBP–USD, EUR–USD, GBP–EUR, GBP–CHF, EUR–JPY, and JPY–NZD. Each observation has a millisecond-accurate timestamp and an identified trading entity, which is almost always a named bank (for example, Barclays or Santander). In addition, there are bid and ask quotes given by the trading entity. The term 'trading entity' is deliberately used to reflect that, within a bank, many individuals or business units may be trading in the interbank currency market simultaneously. Each currency pair likely has slightly different characteristics, and the number of observations for each pair over the day varies. Table 3.2 provided later in the analysis illustrates this idea and shows the differences in the levels of HFT activity that can be detected.

Ersan and Ekinci (2016) and Ekinci and Ersan (2018) also work on the principle that HF traders have a strong ability to update or withdraw orders before they are executed. Unlike a purely OTR-based approach, these authors have access to data from the Istanbul Stock Exchange (Turkey). This dataset consists of five months of data-capturing message flow to the exchange's matching engine. The dataset includes trader IDs, order sizes, prices, etc. This allows the chronology of an individual trader's order flow to be examined and time intervals to be captured, as well. In addition, access to a dataset that is not freely available allows for this approach, and for using an approach that is available only to select researchers or the trading venues themselves.

The basis of the concept is that HF traders can be distinguished by quick reactions. The researchers believe that if a trader cancels, updates, or amends an order in the same direction, within one second of an initial order submission, the trader concerned is undertaking HFT activity. Any algorithmic trader with significantly low latency can meet this condition. Ekinci and Ersan (2018) show that, within their sample of Istanbul-listed stocks, 1.23% of activity triggered this threshold, and 50% of the activity was observed to relate to around only 10 out of a total of 422 equity listings.

3.4.3 Detecting high-frequency traders and activity

Detecting HFT activity is a challenge for researchers and regulators, as markets represent an amalgamation of individual activities. Many datasets are considered proprietary and are not made available. The methodologies available in the relevant literature are summarised below, and these vary in many respects. It is important to note that these methodologies do not necessarily reflect the wider tools to which an exchange venue can access.

It is known that characteristics that are not directly related to latency can be indicative of HFT activity; however, the ability to detect latency is limited because the basis on which a trade is made can never be known. This is the entire basis of how information moves through markets!

The simplest method is a direct method whereby the purpose of a trader's activity is known or their use of latency-reducing systems is known. In these cases, a definitive identification can be made. The direct method is seldom accurate as it can lead to HFT activity being overlooked if an institution's primary activities are not based on HFT or if rapid activity is rare and coincidental (Bouvert et al., 2014). This method simply involves identifying (flagging) trades that represent an update or resubmission time below a certain value (a typical value is 100 milliseconds). Merit can be found in passive identification as described by Bouvert et al. (2014);[14] however, measures based on volume and inventory holding characteristics as well as on the rate at which

[14] This is a paper from the European Securities and Markets Authority.

orders are modified or cancelled can also be helpful, as they allow for a more active screening method (Stenfors and Susai, 2019). It is also possible to consider the individual bid-ask spreads, as it is generally accepted that HFT activity does lead to the tightening of bid-ask spreads when the HFT activity is market making (Petrella, 2006; EUREX, 2013). It is important to note that variation in bid-ask spread is a relative predictor and cannot be considered a substitute for measurements of latency in totality. However, Bouvert et al. (2014) describes and indeed is compelled to use a simpler flag method where datasets do not allow individual orders to be identified, but rather, just the trader is identified. Here, the time gap between one submission and the next from an individual trader is calculated and flagged if this time gap is below a certain threshold – perhaps a tenth of a second.

The method proposed by Stenfors and Susai (2019) is based on having more information in the dataset than the method demonstrated by Bouvert et al. (2014). The dataset in Stenfors and Susai (2019) was obtained from EBS and is no longer accessible to academics. It includes the unique trader identifications, accurate timestamps, details of the order types, and markers to indicate the update of an order. This allowed for screening according to order type and update times.

The goal was not to measure HFT as such but rather to consider the speed at which liquidity is provided and the duration for which it is made available. Similar data was used in the study of Moore and Payne (2011), whose dataset is sourced also from EBS, although the data period is a few days in 1999 in the case of Moore and Payne's study. The focus was not HFT activity; rather, the researchers made use of highly accurate synchronised timestamps to examine private information use in ForEx markets.

An even more extensive dataset was used in an occasional paper published by the FCA (the UK's financial conduct regulator) (Aquilina and Ysusi, 2019). The dataset comprised one year's worth of order book data (2013) with a sample of 60 equities from the FTSE 100 and 60 equities from the FTSE 250. Data was collected from the LSE, BATS, and Chi-X exchanges. Although these researchers had access to an extensive dataset, they opted to use screening methods based on the ways in which individual firms operate and how they use their technology. As such, these researchers do not simply state that quick reaction times alone denote an HFT. Ergo, the FCA researchers use knowledge of individual firms to flag their datasets for HFT activity (Aquilina and Ysusi, 2019).

A rather different method to identify HFT activity uses inverse reinforcement learning (Yang et al., 2012). This methodology seeks to understand the reward function of a trading algorithm to allow a reverse understanding of its decision-making characteristics. This methodology is likely to be restricted, as identifying the deterministic behaviour of a function is likely to be difficult at best, especially if a trader is subjected to a randomisation mechanism or if a trader runs multiple algorithms. This renders the development of a Markov switching model more complicated as it requires a greater number of states to exist. An additional limitation is understanding the trading strategy used (for instance, market making or spoofing[15]). By understanding the neural flow process through backward induction, the order generation time can be calculated, and Yang et al. (2012) estimate this to average at 0.35 seconds for an HFT compared to 20 seconds for a market maker and 120 seconds for an opportunistic trader.

The dataset in this study is sourced from Tick Market Data (Paris) and is originally collected by Reuters FX, this has high frequency timestamps and trader ID values which allow for latency-based analysis. The limitation in this dataset is the lack of known high frequency traders, the parties involved are mostly investment and commercial banks.

Based on the idea of reaction times within the trader's order flow being a method to detect HFT activity, the present researcher screens the dataset in a similar way. To begin a screening of individual trading entities, tick gap times are calculated and a flag is attached to any tick gap of less than one second. As the dataset contains five different currency pairs and the number of traders in the set varies, the researcher presents the values separately for each pair in Table 3.2 below.

[15] This is a candidate method for identifying spoofing, which is a vexatious strategy and is illegal in many countries.

Table 3.2 Estimates of HFT activity using one-second flags

Currency pair	**Number of trading identities**	**% of activity with an internal update time of less than one second**	**Number of trading identities involved**
EUR-JPY	31	3.6598%	18
EUR-USD	46	8.7349%	23
GBP-CHF	7	6.5144%	7
GBP-USD	40	8.1926%	24
NZD-JPY	4	3.5424%	4

As shown above, the activity taking place in less than one second varies quite considerably, suggesting that each market has its own characteristics and there is no normal level of HF activity across multiple ForEx markets.

As one second is quite a long time in the sense that it is known that HF traders are capable of operating at much lower latencies, a second attempt is now presented to illustrate the proportion flagged when looking for activity in 100 milliseconds (0.1 seconds). However, as HFT is a relative term and latencies vary across markets, it is interesting to look at rates of activity when the threshold for flagging is varied.

This is comparable with other sources and is less debatable in terms of ambiguity in what exactly the term HFT means. It is not possible to screen the data by order type or follow individual orders to get exact 'update generation times', as used by Stenfors and Susai (2019).

Table 3.3 Estimates of HFT activity using 100-millisecond flags

Currency pair	Number of trading identities	% of activity with an internal update time of less than 100 milliseconds	Number of trading identities involved
EUR-JPY	31	0.1062%	17
EUR-USD	46	0.2353%	23
GBP-CHF	7	0.07897%	6
GBP-USD	40	0.06104%	24
NZD-JPY	4	0.4237%	3

It is clear that the rates of 100-millisecond activity are much lower than the rates of one-second activity. It is also interesting that these values are much lower than the values suggested by O'Hara (2014). This finding shows that HFT is a very relative term and that different marketplaces see differing intensities of HFT activity. Equity markets are generally believed to see higher values; however, currency markets are harder to estimate accurately as most datasets are not complete, as currency markets are largely over-the-counter markets.

The method applied above is the general approach set out in Bouvert et al. (2014) and is used by the European Securities and Markets Authority (ESMA). It is likely that regulators would consider using this approach when they have a clear threshold for algorithmic trading to be considered HFT, as it can be applied retrospectively or in near real time if needed. However, there is a caveat: When using this method, it is not possible to determine why orders are submitted or to track the open/update/close/withdrawal of individual orders. This limitation makes it hard to determine traders' motives, and, when a trading entity is flagged, the question is sometimes asked: Is the entity a HF trader or are the two orders being submitted through two different processing streams? This a valid question, because, in one case, a trader with approximately 21,000 trades in the dataset was flagged only three times. It is possible that this method could over-estimate HFT activity in this way. Bouvert et al. (2014) used a similar method and also cautions of the risk of over-estimation. In the data above, the number of traders involved includes all traders who were flagged.

No discretion was applied given its subjective nature, but some users could do so in order to help them identify those who trade consistently at a low latency.

As the analysis illustrates, the proportion of HF activity is low. Whilst it is shown that fast trading is possible any analysis would at best take this data as a proxy of HF activity which can be used to evaluate models of analysing market microstructure. The following methods of retrospective analysis are demonstrated with this data, based upon this acknowledgment.

3.5 The applicability of machine learning methods

This section evaluates two machine learning methods that are both well established. The rationale is to test the ability of these methods to propose that certain trading activities are HF by nature. This consideration is important to the research area as it relates to the ability to automate supervision and possibly allow for more supervision of markets closer to real time rather than by ex-post analysis.

The intention of this section is to show that these methods are not particularly helpful in terms of identifying a HF trader, partly because HFT is not definable in a temporal sense and it is not distinct in nature from relatively slower activity. It is the reseacher's hope that this discussion will illustrate that, while machine learning may offer many possibilities to regulators, this level of dataset offers little prospect of making this happen. However, regulators may benefit from AI, conditional on data availability (Butler and O'Brien, 2019).

Machine learning methods are forms of computational AI that react to information and make decisions based on instruction or prior experience (Macchiarulo, 2018). Broadly speaking, there are two forms of machine learning: supervised learning and unsupervised learning. In supervised learning, an application is given training data, and the system is able to refine its ability in classification. In unsupervised allocations, no training data is used, and it is unlikely that the system can become more accurate over time. Supervised learning has its advantages as it allows for better handling of noisy data, and a process of cross-validation can be used to test how the system will react to noisy data in advance (Macchiarulo, 2018).

3.5.1 K-means clustering

The first candidate method is the seminal K-means clustering algorithm, used to extract classification information. This method allows for the use of clustering to derive characteristics of the demographic of the marketplace. In this latent variable model, the aim is to conduct a profile analysis whereby the latent variable is categorical; however, the other variables used are continuous in nature.

The basis of the analysis is the K-means clustering algorithm, which is designed to divide a number of observations (n), into a set number of clusters (K). This is an iterative process whereby means are set and observations are grouped around the closest mean. The algorithm then revises the position of the mean, and the process repeats until a set of groupings with the lowest variance around the means is achieved (Hartigan and Wong, 1979).

The application demonstrated below is minimising the squared Euclidean distances (R) between observations and cluster centroids. It is expressed as a formula where C_k denotes a cluster, x_i an observation, and μ_i a candidate centroid:

Equation 22

$$R(Ck) = \sum_{x \in Ck} (x_i - \mu_i)^2$$

The cluster centroid is in the position of the arithmetic mean of the observations with the cluster. It is expected that a well-defined cluster should have a lower variance.

The K-means clustering algorithm is an unsupervised machine learning technique. An unsupervised algorithm is only capable of generating the output variable using the input variables, as there is no prior knowledge of the state of the output variable given. In this case, the aim is to classify a datapoint to a best-fit cluster/centroid, given a specification for a number of clusters (k). Specifying the number of centroids is best achieved through the purpose-based approach, whereby known characteristics are used to determine the most relevant number of clusters. The relevance is important as it determines the ability to interpret results produced by the clustering algorithm. When the purpose-based approach is not practical, the elbow method is available. This technique is based on the reduction of the sums of square distances as further clusters are added. The method as described is a visual method and requires an eyeball interpretation of a graph plot. It may be possible to formalise an indicator to make it a more rigorous test.

In order to explore this idea further, the K-means clustering approach is run using the calculated market tick gap and the individual trader's tick gap, as these are two continuous variables. In both cases, lower values reflect lower latency.

One of the key challenges here is setting the appropriate number of clusters for identification. In the first instance, the output is presented with six clusters, as the dataset contains six traders.

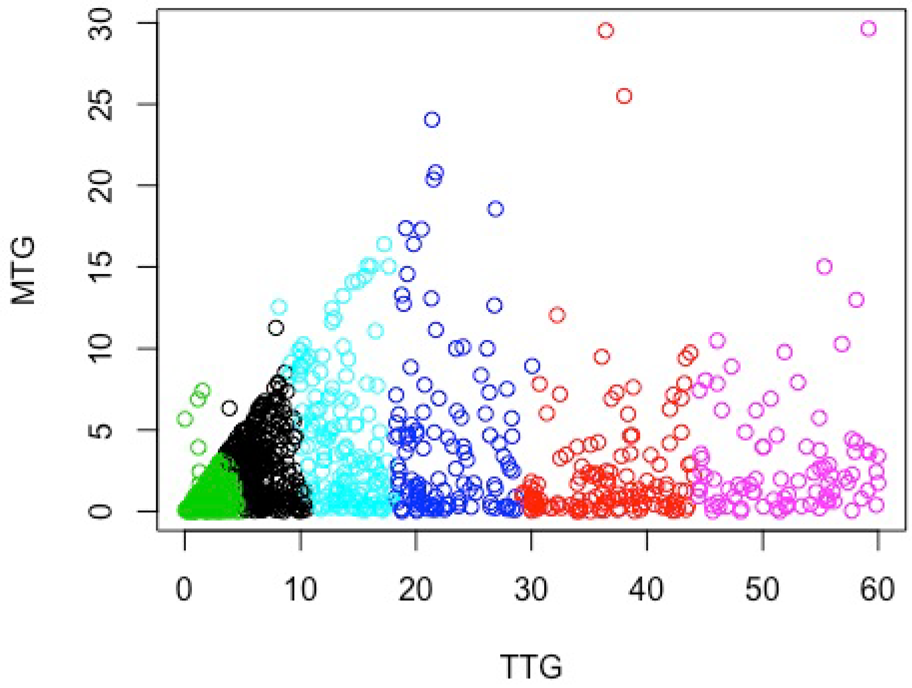


Figure 3.5 K-means cluster membership denoted by colour (k = 6)

The axes are plotted in seconds. MTG denotes the market tick gap and TTG denotes the trader's tick gap.

Each cluster is presented in a separate colour, and each point represents one pair of observations. As is evident, the density varies. What is most important to note is the lack of separation between the clusters, which suggests that one cluster has little meaning when compared to another. It is also important to note that the clusters need not contain an equal number of observations. Here, there are six clusters of sizes 348, 127, 833, 110, 152, and 85 (left to right/green to pink). As HFT activity is not well defined, it could, for example, be debated which clusters do or do not indicate HFT activity.

This chart is helpful as it shows that the majority of activity is undertaken at lower latencies; however, a simple scatter plot would show the same thing.

It is important here not to specify the model with too many clusters. To buttress the observation that this method is not helpful, the researcher now shows the within sums of squares values for iterations with between two and six clusters. There is a commonly applied (although not rigorous) test known as the 'elbow method' that can be used when these sums of squares are plotted. It is said that, when the downward gradient becomes shallower (an elbow shape), this is an indication of the point at which adding additional clusters has diminishing value. It is clear in the chart below that this point is where there are only three clusters.

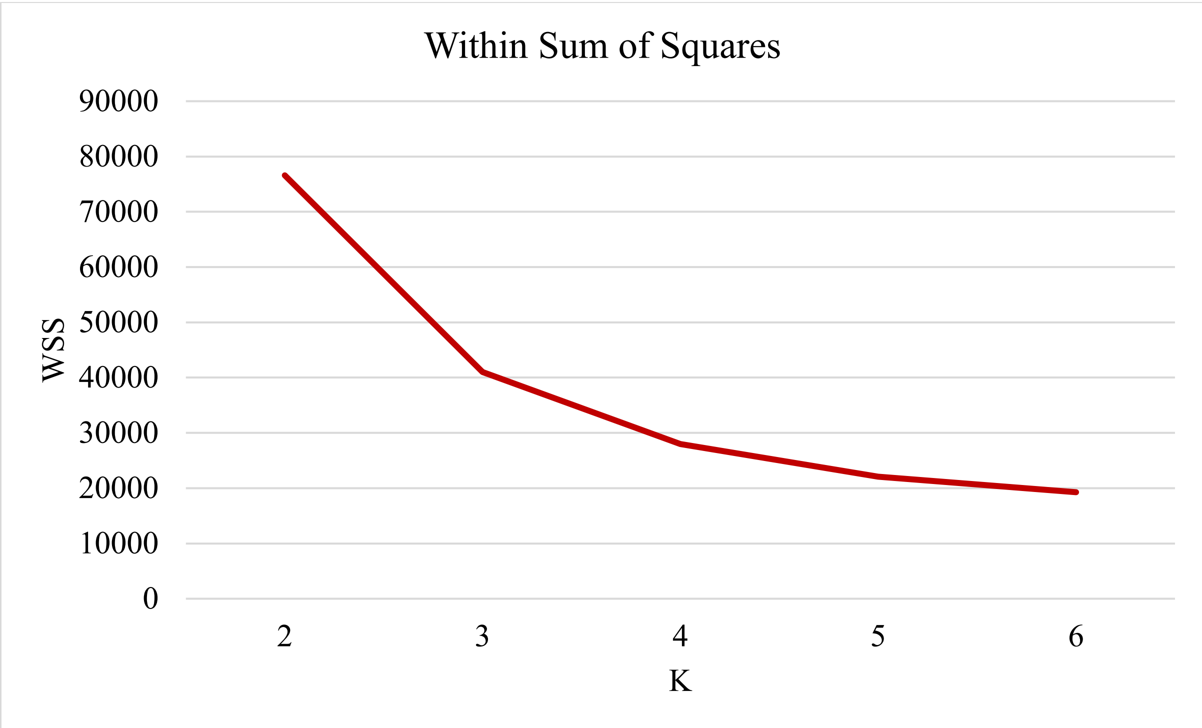


Figure 3.6 Within sums of squares for K-means clustering

Below is a graph that shows how the cluster membership has been allocated when the algorithm is constrained to three clusters. In this graph, the clusters seem to be a little more well defined – as is evident from the point at which the green and red clusters meet. Although this estimation does identify a faster cluster (number one in green), it is not clearly defined to the extent to which it can be concluded that it represents HFT activity. The axes are plotted in seconds. While these are quickly executed trades, it may be pushing the generally accepted meaning of the words 'high frequency' a little too far. For this reason, this method cannot be considered a robust HFT detection method.

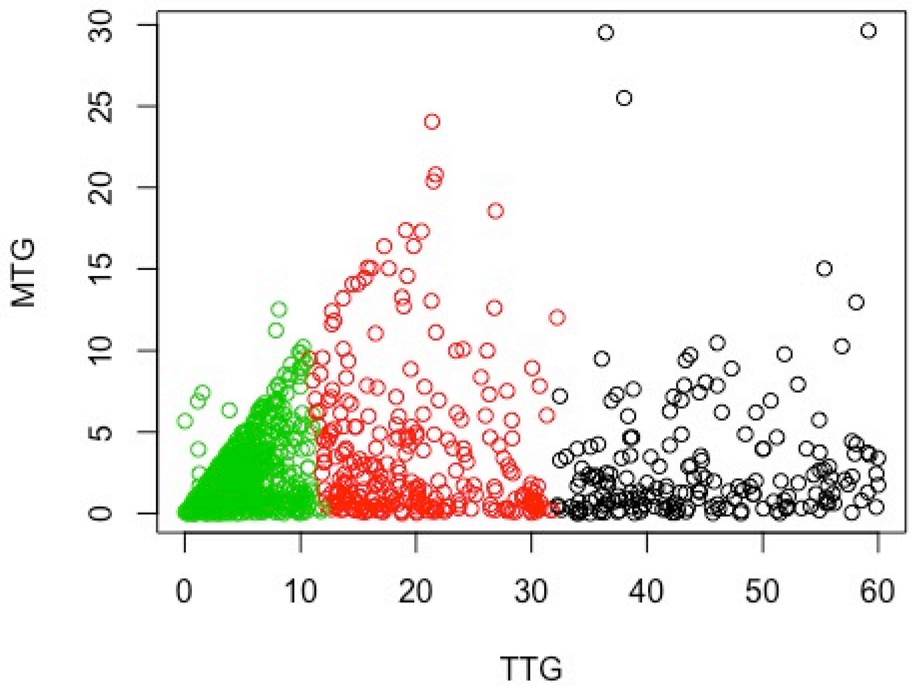


Figure 3.7* K-*means cluster membership denoted* by colour (k = 3*) – axes are plotted in seconds

3.5.2 Factor classification

Classification tree methods are perhaps best reserved for models wherein the classification can be obtained by no other means (Magidson and Vermunt, 2002; Wang et al., 2013). When used in unsupervised applications, the accuracy of the results may be very poor in the attempt to allocate a latent variable that is based on relationships between observed variables (Mourad et al., 2013). The positioning of the variables within the model depends on the directed or undirected nature of the network; for instance, in a directed network leaf, variables must be observed, which ae captured by a set of edges between nodes (Mourad et al., 2013).

When the exact structure of the network is unknown, it becomes necessary to determine the architecture of the network, which can be achieved by optimality scoring using criteria such as the Bayes information criterion (a goodness-of-fit measure that penalises models with a greater number of variables). Discovering the optimal structure of the classification tree can be computationally challenging, especially when it is not initially obvious how variables relate to one another. In such a case, it may be possible to add latent variables to the model in order to aid classification by means of non-observed indicators (Mourad et al., 2013); unfortunately, this is not helpful in the case of the HFT dataset, especially as market composition and structure may vary over time.

Magidson and Vermunt (2002) compared two methods and found that the supervised approach produced a 1.3% misclassification rate, while the unsupervised approach produced 1.7% misclassification.

Although accuracy is an issue, and it is possible to have a much higher error rate than the source suggests, this approach does not require the setting of a cluster value (k). In this case, in which a semi-supervised approach can be used, it is not necessary at all.

The dataset only contains one variable: trader ID. The factor classification method uses continuous numeric variables that explain the classification of an observation. For example, taking various measurements of the petals of a flower would allow a researcher to classify the flower by species.

In order to explore the dataset further, the latency variables MTG and TTG (as defined earlier) are used to try and predict the originating trader. Figure 3.8 below shows this process:

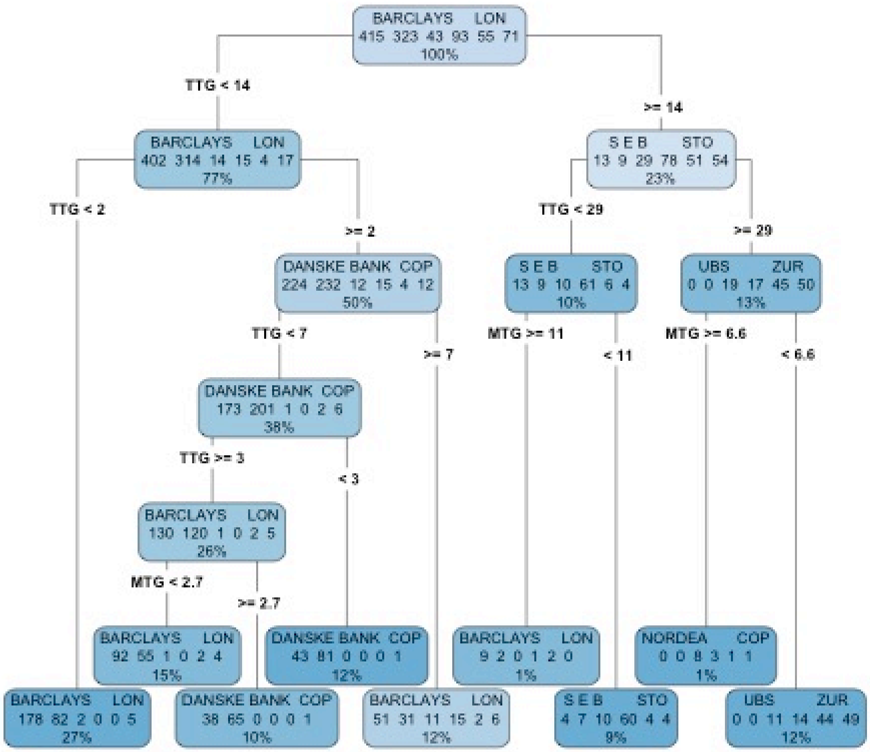


Figure 3.8 Factor classification tree – the numbers on the second row in each blue box denote correct and incorrect classifications

If this were a successful application that identified one or more particular trading entities as HF traders, then it would be clear that the groupings at the bottom of the figure are accurate. An accurate grouping would have one large number and all others at/close to zero. In fact, what is clear in this example is that the use of the trader tick gap and market tick gap together are not attributes that allow us to work backwards to identify an original trader.

This is not a successful application of a factor classification model, as many traders were misclassified. This finding shows that the latency behaviour of an individual trader is not predictable to the extent that using this process alone cannot describe an originating trader. In short, using only the information in the dataset as a given, the classification tree cannot be used to reliably classify one trader over another as an HF trader. The possibility of training the AI, using data wherein HF traders are indicated, could help identify HF traders in later data in which HF traders are not identified. This method would allow us to avoid relying on finding the fastest proportion, as in the K-means example.

The two machine learning methods demonstrated have not provided a useful way to analyse the data; however, this line of enquiry is worthwhile as it could be a very useful way of monitoring and analysing datasets. Furthermore, if this had been a viable method, it would have been the first demonstration of HFT detection using an existing machine learning method.

3.6 The proposed method of examination

3.6.1 Price and its explanatory effects on bid-ask spread

The relationship between price and bid-ask spread can be shown as positive: as prices rise, the bid-ask spread tends to widen (Narayan et al., 2015). However, it is possible in this research to capture the sign, size, and significance of a bid-ask spread coefficient. This may either serve as confirmation of the finding in Narayan et al. (2015) or provide an interesting contradiction.

In the context of markets that may be characterised by HFT, it may be possible to capture traders' reactions to changes in the bid-ask spread. A priori, based on the findings in Petrella (2006) and what is known about scalping strategies, it is likely that, as the bid-ask spread narrows, the possibility of scalping becomes more viable given that its aim is to close small spreads. The basis of a scalping strategy is to hold an asset for only a very short period of time and remain inventory neutral; hence, HF traders are rather well suited to scalping strategies. Silber (1984) defined scalping as frequent trading in small quantities while maintaining an inventory-neutral position. Petrella (2006) argued that, in options markets, scalping is likely to take place when the bid-ask spread is tighter. Using the ForEx dataset in this chapter, it can be shown that HFT activity increases when the bid-ask spread decreases; ergo, it is probable that the HFT activity identified in the dataset does, at least partly, reflect traders' running scalping strategies in an over-the-counter market.

In order to explore this point, a dummy dependant model can be used to test the directionality of the probability that traders with higher bid-ask spreads will be flagged as HFT activity. The dataset used is the EUR–USD currency pair, which is described earlier in this chapter. This dataset was found to contain 0.2353% HFT activity when analysed using the flag approach with a threshold of 100 milliseconds. This currency pair has a competitively high incidence of HFT activity; therefore, it provides better scope for a robust finding.

The model used here is a dummy dependant model, whereby the dependant variable either takes the value of zero or one. In this case, the number 1 denotes the presence of an HFT flag. The estimated equation is:

Equation 23

$$HFT_{flag\ i} = \alpha_i + \beta BAS_i + \varepsilon_i$$

Estimation of this linear probability model by means of ordinary least squares produces the following estimation output.

Table 3.4 Estimation output indicating the probability of an HFT flag

Variable	EUR–USD OLS coefficient
Intercept	0.541
Bid-ask spread	-0.359 ***
	(0.008)
N	1048575
R-Squared	0.002

*,**,*** *denote significance at the 10%, 5%, and 1% levels, respectively*

Standard errors in parentheses

This model takes the form of a linear probability model, where the interpretation of the coefficient of the bid-ask spread is in probability terms – in this case, the probability that the HFT flag will be applied (indicator equals 1). The coefficient, -0.3598, represents the probability that an increase in the bid-ask spread equal to 1 will lead to a decrease in the probability (of 0.3598) that the HFT flag will be applied to a trade.

As this is a linear probability model, a few issues arise. In particular, it is not possible to interpret the R-square, as not all of the variables are continuous in nature. The non-homoscedastic distribution of the error term is also noteworthy. However, the *t*-

statistics do remain valid and in this case are statistically significant at the 1% significance level.

This estimation output illustrates that the bid-ask spread does have an explanatory effect on the flagging of HFT activity, which more importantly suggests that the dataset contains algorithmic traders (who meet the HFT threshold) who are using strategies such as scalping. While the dataset does not contain information about order types, this finding indicates that both HFT latencies and HFT activities are captured in the dataset.

3.7 Outline of the method

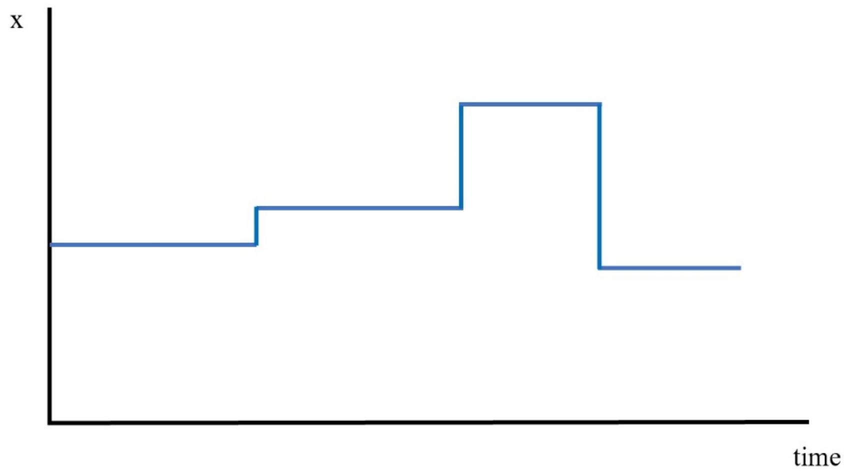
Spot prices reflect all readily available information for all *t*. These are given by a fair-game process:

Equation 24

$$x_{it} = \Delta x_{it} + \sum_{i=1}^{t-1} x_{it}$$

Here, x_{it} is not exactly the spot price; rather, it is a working price level. It is necessary to calculate this parameter as it reveals an individual trader's perceived correct price, even if they have not traded at a given time *t*. This allows a working price level to exist for every trader (*i*) for all periods (*t*).

Ergo, a change in bid or ask will 'jump' the price level, creating a time series that looks like this example:



A fictitious example of working price levels moving over time.

Figure 3.9 Price working levels vary over time periods

This series likely does not have a normal distribution if prices are random walk with drift. The crux here is the uneven distribution of Δx_{it} in Equation 9 creates the heteroscedasticity problem. Hence, an autoregressive conditional heteroscedasticity (ARCH) or generalised autoregressive conditional heteroscedasticity (GARCH) is needed.

Here, the AR component regresses residuals on lagged residuals, and the CH element tests the extent to which prices are conditional on previous prices. As a general principle, an ARCH order higher than three would suggest that a GARCH approach may be more suitable.

The following is an ARCH (1) structure:

Equation 25

$$\sigma_t^2 = \lambda_0 + +\lambda_2\sigma_{t-1}^2$$

Adding in some explanatory variables gives:

Equation 26

$$\sigma_t^2 = \lambda_0 + \lambda_1\sigma_{t-1}^2 + \lambda_2 P_{Trader_t} + \cdots + \varepsilon_t$$

σ_t^2, on the left-hand side in 25 and equation 26, denotes:

Equation 27

$$\sigma_t^2 = Var\ (\varepsilon_t) = h_t$$

Let h_t be the variance in market mid-price over a period $t(1, \ldots, T)$. This can be reduced to a sensible period on a rolling regression basis. It is also possible to give h_t by the following estimation, which illustrates that h_t and ε_t are one and the same:

Equation 28

$$P_t^{MID} = P_{t-1}^{MID} + \varepsilon_t \quad \sim IID(0, \sigma_t^2)$$

Thus far, it is assumed that variation in price can be explained by past prices, passage of time, and market depth. These are the only explanatory variables available in the tick market dataset (PAR). This is reflected in Equation 26. It is worthwhile to examine the dataset in full; however, it is extremely likely that other variables would be explanatory, as well.

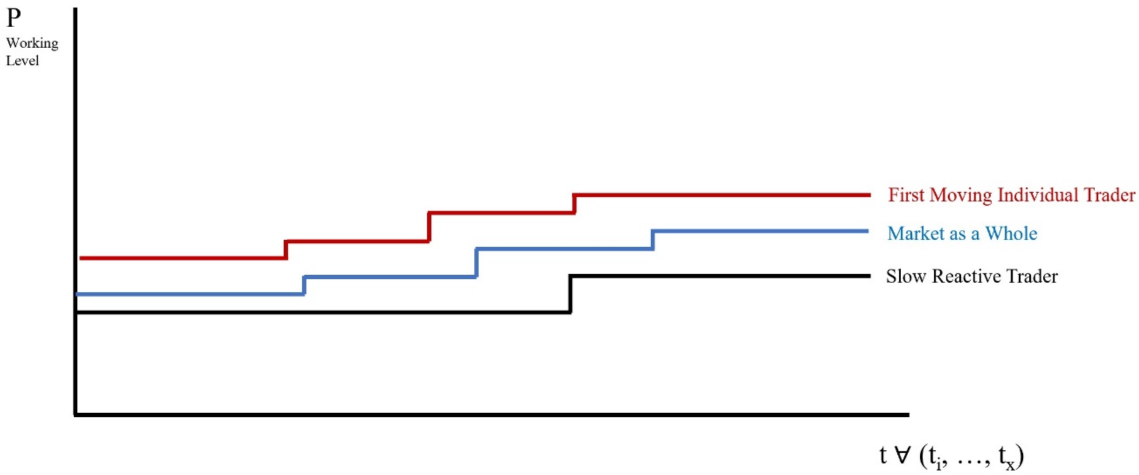


Figure 3.10 The market average working level, shown with slower and faster traders

Here:

Equation 29

$$\bar{P}_i = \sum_{t,\ldots,T}^{i=1} P_i^{LEVEL} \quad \forall t$$

ΔP here is:

Equation 30

$$\bar{P}_{i,t} - \bar{P}_{i,t-1}$$

The next section summarises the characteristics of the dataset and then demonstrates the proposed method using the dataset, where each trader is given a working price level series of their own.

3.8 Data analysis

3.8.1 Dataset description

The estimations that follow were made based on a dataset obtained from Tick Data Market, a data provider based in Paris that provides datasets on a commercial basis. This data captures part of the interbank currencies market, which comprises two intermediaries (NEX/CME and Reuters). This dataset uses data that originated from within the Reuters element of the interbank market; this is not a HFT venue. This dataset is not the full interbank market and in no way could be said to represent the entire ForEx market. Again, this is at best a proxy of HFT activity, whilst the dataset illustrates some fast trading, the venue from which the data is captures is not believed to allow HFT strategies and the banks involved would not be likely to deploy these strategies anyway.

The dataset is taken from Wednesday, 01st June to Friday 03rd of Jane 2016 and covers the whole trading period. Six separate currency pairs are included: GBP–USD, EUR–USD, GBP–EUR, GBP–CHF, EUR–JPY, and JPY–NZD. Each observation has a millisecond-accurate timestamp and an identified trading entity, which is almost always a named bank (for example, Barclays or Santander). In addition, there are bid and ask quotes given by the trading entity. The term ‘trading entity’ is deliberately used to reflect that, within a bank, many individuals or business units may be trading in the interbank currency market simultaneously. Each currency pair likely has slightly different characteristics, and the number of observations for each pair over the day varies. Table 3.2 provided previously illustrates this idea and shows the differences in the levels of HFT activity that can be detected.

As the datasets for all currency pairs do not all contain the same trading entities, it is best to treat them as separate datasets; hence, all estimations are presented separately.

3.8.2 Dataset analysis

The working price levels, as set out above, have been calculated for every trading entity. For larger sample sizes and where traders are frequent, these series are rather indistinct from ordinary price series. The example below is taken for one trader from the GBP–USD currency pair and plots the overnight period between 02 and 03 June 2016. This graph illustrates the volatility that is retained in the data but also shows that, when plotted, horizontals appear to show that prices have not moved.

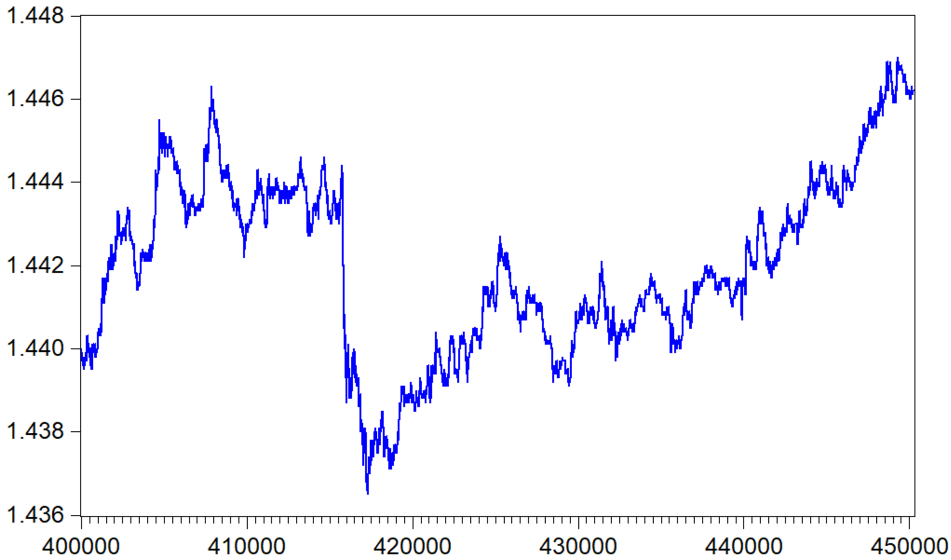


Figure 3.11 Working price level for Barclays for an overnight period (02–03 June 2016) while trading the GBP–USD currency pair. Note that the x axis shows the observation numbers – taken from approximately the middle of the time period

Table 3.5 Estimation outputs for currency pairs at working levels within the variance equation

		ARCH Estimation Order 1			
Pair	GBP USD	GBP CHF	EUR USD	EUR JPY	NZDJPY
		Reference Equation			
Intercept	0.0000***	0.0000***	0.0000***	0.0085***	-0.00054
Mid-Price(-1)	0.9999***	0.9999***	0.9999***	0.9999***	1.0009***
		Variance Equation			
Intercept	0.0000***	0.0000***	0.0000***	0.0001***	0.0004***
Residual(-1)^2	0.3482***	0.1714***	0.1500***	0.1500***	0.4427***
Barcalys	0.7823***	0.7733***	0.9061***	0.7788***	
BTM	0.0000		-0.0006***	0.0030***	
Cibc				-0.0011***	
Citi Bank			-0.0004***		
Comm Bank	-0.0005*		0.0019***		
Danske	0.0641***	0.2259***	0.0009***	0.1407***	
DNB Bank	0.1342***		0.0006***		
HSBC	-0.0048***		0.0758***		
Kaspi Bank	-0.0005**				
Nordea	0.0028*	-0.0001			
Piraeus Sofia				0.0027***	
RBS	0.0011***	0.0000			
Sabadell				-0.0010***	
Santander	-0.0003***				
SEA				-0.0026***	
SEB	-0.0002***	0.0007***		0.0277***	
Soc Gen	-0.0006***		0.0021***	0.0014***	
Swiss Fin	-0.0003***	0.0000	0.0000	0.0029***	0.0000***
TKFX				-0.0012***	
UBS	0.0218***	0.0002***	0.0077***	0.0491***	
Westpac	0.0003***			-0.0003***	0.0000***
Zuercher			0.0080***		

*,**,*** Denotes significance at the 10%, 5% and 1% level

Table 3.5 shows the same model applied to each of the five currency pairs included in the dataset. Here, the output reflects two estimated equations for each of the five currency pairs. The first is the reference equation, which is an auto-regressive equation of order one (AR(1)). The second equation (the variance equation) has the error term of the first equation as its explained variable and examines the relationship between the working price levels of the trading entities listed.

Gaps in the table reflect that not all trading entities are active in trading each pair. Across multiple currency pairs, some common characteristics are seen in the coefficients of individual trading entities. For instance, Barclays always has a coefficient that is positive and statistically significant up to the 99% confidence level, ergo we are shown that this trading entity has a consistent impact upon the price. It would be logical to ask 'could we think this evidence of HFT activity'? On the one hand, this could be evidence of information being traded into the price or some elements of momentum ignition, yet on the other hand as we do not know the exact reasoning for this a more conservative answer is needed. This is: whilst this is supporting evidence this alone does not indicate HFT activity.

Another interesting case is 'Swiss Fin' which has a very different outward appearance. Here it seems the trading entity has very little effect as the coefficients are very close to absolute zero, indicating no overall effect either way. HSBC is also worth examination as it has a negative and significant coefficient when trading in the Dollar-Sterling currency market. This is indicative of average activity which is trading against the movement of the markets mid-price, either resisting movement or adding noise[16]. Trading against price could be seen as a slow reaction to market conditions, but this could also be seen as a deliberate act to deploy a scalping strategy. These are two scenarios here, the former shows evidence that this is not HFT activity however the latter does indicate this. Again this stresses that this method might offer support to an existing idea and interrogate what activity is taking place but alone can not be used to screen for HFT activity.

An LM heteroscedasticity test is also conducted for each pair separately in order to confirm the presence of the non-normally distributed error term of the reference equation. For the estimated variance equations to be valid and logically interpretable, the ARCH effect must be present. The results of the GARCH Lagrange

[16] The word noise is used in the same terms as Black (1986).

multiplier (GARCH LM) test are provided in the next sub-section. A few notes follow analysing each currency pair in turn.

GBP–USD currency pair

When examining this currency pair, 15 trading entities could be included, as they traded sufficiently consistently across the time period. Thirteen trading entities were removed, as they trade only sporadically. That is, they trade less frequently than once a day in only short bursts.

The coefficient of the auto-regressive element (within the reference equation) shows the extent to which previous prices determine current prices. This value is below one but is close to it, indicating that past prices are highly influential. This coefficient is also statistically significant at 1%.

The ARCH term, which is the square of the lagged residual (labelled as Residual(-1)^2) is illustrative of the heteroscedasticity within the error term of the reference equation. This finding shows the significance of the pervious values of the error term in explaining the current value of the error term, which confirms that previous volatility may explain current volatility in the impulse term. The significance here confirms the presence of the ARCH effect.

The relationship between the market mid-price impulse term (the error term of the reference equation) and the regressions capturing the trading entities' working price levels is statistically significant in most cases. These are the coefficients in which this study is principally interested. The exception is BTM. Furthermore, Comm Bank and Nodera are significant at only the 10% significance level. In the case that one of these variables are insignificant, the interpretation is that the particular working price levels are not statistically different from the market's mid-price (which reflects all traders). This conclusion also suggests that the particular trading entity does not make a significant impact on market price.

GBP–CHF currency pair

The interpretation is the same as that which was applied to the first estimation. This estimation captures all trading entities within the dataset (seven in total). In this case, each of the traders is active throughout the time span and therefore can be included in the model. As shown above, three of the trading entities (Nordea, RBS, and Swiss Fin) do not differ significantly to the market mid-price and do not appear to contribute to the price-making process. Those entities that have significant coefficients (Barclays, Danske, SEB, and UBS) all possess coefficients with a positive sign. This means that these trading entities do impact price, as their working price levels move ahead of the market's mid-price.

EUR–USD currency pair

This dataset contains records of 36 trading entities, 25 of which were insufficient for generating a working price-level series, which leaves 11 trading entities that traded across the day. All but one of the trading entities prove statistically significant. The significant variable coefficients take a mix of positive and negative signs. The finding is that Barclays, Com Bank, Danske, DNB Bank, HSBC, Soc Gen, Swiss Fin, UBS, and Zuercher trade ahead of the market mid-price and contribute to the price-formation process. On the other hand, BTM and Citi Bank trade behind the market mid-price contribute to the price-making process but trade behind the market mid-price.

EUR–JPY currency pair

Of a possible 27, this estimation includes 13 trading entities. Exclusion is based on the previously stated reasons. All variables included are statistically significant at both 5% and 1%. Once again, the signs of the coefficients are mixed, suggesting that some trade ahead of the market mid-price and others work behind it. The finding is similar to those already set out above.

NZD–JPY currency pair

This is the smallest of the five datasets, both in terms of the total number of observations (61,393) and of the number of trading entities within the dataset (four). Only two traded regularly (Swiss and Westpac), and the remaining two entities traded only sporadically and not often enough to produce a working price-level variable for their entity.

The mid- price reflects the market as a whole, and the variables Swiss and Westpac are each statistically significant at 1%. Swiss operated behind the market mid-price (indicated by the positive coefficient) whereas Westpac operated ahead of the market mid-price, on average, as indicated by the negative sign of the coefficient.

In the five estimations, it is possible to use the ARCH terms to examine a trader's contribution to the price-formation process in terms of its ability to contribute to market price. As this method does not use a difference-in-differences approach, it is also possible to also identify which traders, on average, traded against the primary movement in market price and were (or were not) significant in their effect on the market price.

Subsequent sections prove that the ARCH estimation is required due to the presence of the heteroscedasticity phenomenon.

3.8.3 The ARCH LM test

In order to validate any ARCH model, it is necessary to confirm that the model contains ARCH effects. The LM test for heteroscedasticity can be used to confirm the presence of ARCH effects. The LM statistic is compared to a chi-squared distribution: When the null hypothesis is rejected, it is concluded that the model exhibits ARCH effects.

When undertaking this check for each currency pair in the dataset, the following results were obtained.

Table 3.6 ARCH LM test results

Currency pair	LM statistic	P-value	Reject H0?
EUR–JPY	1157.05	0.00000	No – ARCH effects indicated
NZD–JPY	20.96	0.00000	No – ARCH effects indicated
GBP–CHF	210.35	0.00000	No – ARCH effects indicated
EUR–USD	2080.14	0.00000	No – ARCH effects indicated
GBP–USD	2294.03	0.0000	No – ARCH effects indicated

It is also possible to see how the error term in an ARCH model is distributed compared to a normal distribution. For example, the regression error term is plotted below. The distribution of the error term is shown in blue relative to a normal distribution, shown in red. The distributions are shown in quantile terms against one another; therefore, the variation of the residual plot (blue) around the linear red line reflects the non-normal distribution of the error term. What is interesting in this example is that the divergence from normal occurs towards both ends of the distribution and has a double inflexion at the centre, which suggests that the bulk of the observations of the residual are in fact distributed away from the centre of the distribution.

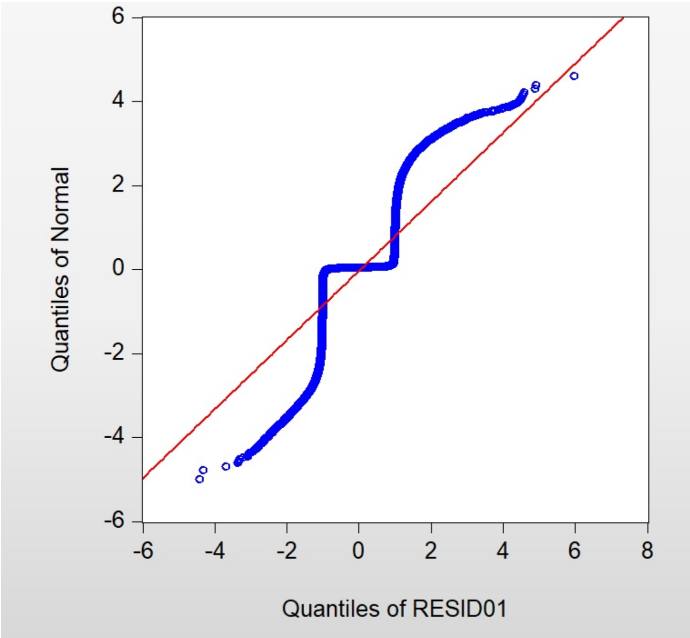


Figure 3.12 EUR–JPY ARCH estimation residuals (blue) compared to a normal distribution (red)

3.9 Working in shorter time periods

In the estimations set out above the time period used has been rather long in an effort to keep the number of observations high. However, this led to a number of trading entities not being included due to their infrequent activity across the period. Therefore, it would be interesting to illustrate that the same method can apply to shorter time periods.

From a regulator's perspective, when analysing traders' behaviours around a known event, a shorter time period may be more helpful for determining how traders reacted to the event.

In order to demonstrate this position, a two-hour period from the middle of the trading day, when prices are quite stable, is taken as an example. The dataset used is the EUR–USD currency pair, as this contains more HFT activity than the other datasets and has a greater number of traders who are active over short time periods. The period of two hours on 01 June 2016 between 10:00 and 12:00 (based on Paris timestamps) was selected at random and not in response to an event or cluster of activity. Once again, the presence of the heteroscedasticity phenomenon is confirmed initially by running an LM test, which shows a 99% confidence level that heteroscedasticity is detected (F_{stat} = 2080.14).

The estimation output below is to be interpreted in the same way as those set out above. This is an illustration that a shorter time period does not affect the viability of the method. Twelve of the working price levels are not statistically significant, suggesting they have no significant ability to contribute to the price-formation process.

Table 3.7 EUR–USD currency pair over a shorter time period

EUR–USD	
ARCH estimation order 1	
Reference equation	
Variable	Coefficient
Intercept	1.119809***
Mid-price(-1)	0.0000985***
Variance equation	
Intercept	0.00000000423***
Residual(-1)^2	0.963186***
Bank BPH	-0.007056
Barclays	0.417969***
BNY Mellon	0.000941
Carl Kliem	0.024163***
Citi Bank	-0.054938***
Com Bank	0.135824***
Commerzbank	0.103491***
Danske	0.049493***
HSBC	0.007539
Kaspi Bank	0.022486***
Komerční	0.004207
LBBW	0.002569
Banca Monte dei Paschi	-0.008729
Nordea	-0.006826

OTP Bank	-0.001947
Rabobank	0.190467***
RBS	0.14189***
Soc Gen	-0.034756**
Swiss Fin	-0.026966**
Trinkaus	-0.001955
UBS	-0.018230
Westpac	-0.000900
WGE Bank	0.065945***
Zuercher	-0.007598
N	8907
Period	01 June 2016 10:00–12:00 (ECT)

,,*** denote significance at the 10%, 5%, and 1% levels, respectively*

This estimation is based on 8,907 observations – for only two hours of data. As the density of data varies across time, it is not possible to be precise. Some sample time periods picked at random for the ForEx dataset illustrated that 20 minutes could give 1,107 observations and 60 minutes 3,363 observations.

The time periods that are viable for estimation are constrained by the number of observations needed to estimate the model.

Naturally, as this model does have several explanatory variables, a certain number of observations is required to match the degrees of freedom used. ARCH models require a sufficient number of observations in order to capture any cyclical effects that may be present. For example, Box and Tiao (1975) suggested that, when using monthly data, a researcher should aim to include at least 50 observations. For ARCH and GARCH models, the strength of the auto-regressive (correlating relationship) and the number of observations required to prove it is significant are related. For a correlation of 0.9987, which is approximately what these datasets contain, a minimum of 2,000 observations would be needed (Ng and Lam, 2006). The effects of HF datasets on this minimum requirement are not clear, although Ng and Lam (2006) did use financial data in their likelihood estimation – the data considered daily frequency rather than intra-day frequency.

This finding suggests that there is a minimum time period that can be captured; however, this minimum time period varies according to the density of the data. For example, one 20-minute period contained 1,107 observations, which is insufficient per Ng and Lam (2006). However, extending that 20-minute period to 60 minutes would give 3,363 observations, which should suffice. The estimation of the variance equation also presents a challenge given the number of explanatory regressors this may need to contain. Naturally, degrees of freedom can be preserved by not including variables for trading entities that do not trade in the time period (as the working price value would be constant). To further increase the reliability of the results, variables that prove insignificant in explaining variance in the price impulse could be removed.

To a certain extent, a pre-screening process could be conducted using a simple multiple regression of the trader's working levels against the mid-price; however, this is not likely to be a reliable estimation. It may be worthwhile to consider which variables can be excluded due to the market mid-price being indistinguishable from the trader's working price level; ideally, a near-zero coefficient that is also not significant at 5% and 10% significance levels.

A practical note to anyone attempting to replicate this method is the real possibility that, when a working price level remains constant, there is a chance that it will correlate perfectly to another constant price level. In such cases, a non-variant price level is of no practical interpretation and should be removed from the model. If this method is estimated in an econometrics package, a near-singular matrix error will likely be caused by this problem. Some econometrics packages are sensitive to variables that have no observations (such as EViews); so, when replicating this method for short time periods, caution is needed to ensure only active traders are modelled. This second practical point would likely give an error of insufficient observations if it is neglected.

3.10 Conclusion

The method of finding first and early movers provides users of basic datasets with a tool to investigate datasets that are commercially available to independent researchers without the need to obtain data from trading venues. This chapter has also shown that this method is applicable over shorter timeframes, which allows for the study of activity around an event of interest or over a period of time, which could be as low as 20 minutes.

In the estimations based on the dataset presented above, it is possible to determine which individual traders can influence the price impulse term (those that are statistically significant). Furthermore, those with a negative coefficient can be shown to be reacting to the impulse term more slowly; in some way, they are resistant to price impulse. Ergo, it is possible to identify which traders are engaging in active price formation and might be of interest in explaining a particular event or when screening traders by activity type. Of equal interest is the ability to find that a particular trading entity does not have any part in the price formation process.

This method can be used to indirectly screen for HF traders, as they have been known to affect price and use momentum ignition strategies; ergo, an HF trader would likely have a relatively high and significant price impact coefficient. In general, HFT also aims to trade new information into price as a part of its strategy. An effective HFT would thus be expected, as it incorporates information into price and therefore has a positive coefficient and a statistically significant relationship with the market price. However, this is not an exact screening for HFT, as all trader types have the ability to incorporate information into price.

For example, if the goal is to study a flash crash-type event, this approach would allow a short time period to be analysed. In that analysis, it would be possible to observe which traders contribute to the price formation process (positive coefficients and significance). Similar in interest would be those trading entities who possess negative coefficients, as these would have been the traders resisting the price movement, trying to correct the error and return the market to the correct 'rational' price. The 2010 flash crash is one of the most widely studied market volatility events concerning which it has been possible to decompose the market (Easley et al., 2011; Kirilenko et al., 2017). However, these researchers largely quote from a Securities and Exchange Commission (SEC) report and perform analyses of their own that do not use a trader-by-trader dataset.

This ability to deconstruct the part of the 'black box' of trading allows for independent oversight of markets and particularly algorithmic traders who are rather active. The next chapter of this thesis evaluates the regulations that are in place for governing the design and deployment of algorithms and illustrates how little is done proactively to monitor behaviours in the markets for exchange-traded assets. However, this chapter has shown that it is possible to generate a certain level of intelligence from available data without excessive difficulty. Therefore, it is possible for regulators to monitor markets and examine a wide range of events given that a good deal of effort is omitted by narrowing down which traders are contributing to price; ergo, it is only necessary to determine the basis of a few traders' actions to understand why a market movement took place.

The FCA, like many conduct regulators, takes an interest in active monitoring (FCA, 2018); however, this is a task that would be very hard to achieve in reality, as it requires the ability to analyse a big data set (Seddon and Currie, 2017). As part of a screening process, the proposed method could prove useful for analysing specific market moments as identified by a threshold-sensitive algorithm.

The next chapter in this thesis evaluates the international regulatory framework that applies to HFT and algorithmic trading and demonstrates the strengths and weaknesses of various market intervention strategies covered in the literature. A more in-depth analysis of the big data challenge is presented followed by a discussion of how methods such as that which was proposed in this chapter can be integrated into RegTech solutions.

4 Regulating Low and Mixed Latency Financial Markets

4.1 Abstract

Global financial markets are governed by a collection of national regulatory systems and a limited set of super-national and non-legally binding frameworks. A key limitation thereof is the poor coverage of cross-border activity and arbitrage strategies. In addition, little direct regulation of HFT exists. This chapter outlines the extent of and variations in regulatory approaches and aims to identify areas of good practice in the regulation of algorithmic trading in which traders may seek a latency advantage. On balance, this is a sensitive area for regulators, as some HFT activity can be shown to generate negative externality effects and market asymmetries may be seen as 'unfair'. This chapter evaluates the role of national regulators, the scope for regulatory technology, and the potential for a rules-based (or ordoliberal) regulatory approach.

4.2 The role of financial services

In this chapter, a critique of the regulatory environment is offered. The aim thereof is to illustrate the differences between the regulatory structures of different countries and to show the potential issues that arise when activity takes place across multiple regulatory jurisdictions. This has become an issue, especially in Europe where assets can be listed in multiple venues that are open at the same time and are subject to different regulatory regimes. This cross-regulatory asymmetry allows for instability issues to spill over into other trading venues.

The chapter then surveys the tools available to exchanges and regulators, and the potential effectiveness thereof is assessed in terms of their ability to prevent abusive activity and minimise the potential for market volatility events. A contribution is made by considering the extent to which an intervention could be seen as systematically fair. Of course, 'fairness' is a complex and subjective concept that this study explores by evaluating elements of the existing literature concerning ordoliberalism and the use thereof as a framework for regulation.

In terms of analysing the regulatory structures and the burden of compliance placed on trading firms, two issues are relevant. Firstly, small HFT operators lack the ability to effectively self-regulate. Secondly, large firms likely face the problem of being too big to manage. These problems, along with the general challenge of cross-border activity, give rise to a series of recommendations aimed at both conduct regulators and those who manage trading venues. The recommendations focus on minimum resting times, the limitations of self-regulation, the need for regulators to access data, and minimising the risk of spillovers between markets across national borders.

It is important to begin by considering the role of HFT in the wider economy. Broadly speaking, the method of trading in 'fictitious commodities' likely has little effect on the real economy (Cerpa Vielma et al., 2019). The term 'fictitious commodity' refers to the concept of labour power – the effort exerted to produce an item (Collins, 1984). An asset is produced using workers' labour power to create or enhance the value of capital. Equities, bonds, or derivative instruments do not derive value from themselves; rather, they derive from an underlying entity. For example, the value of an equity derives its value from the profitability or expected profitability and long-term viability of a firm, which in turn is derived from the firm's production of goods or services, which use labour power to enhance the value of capital. Ergo, according to this logic, financial assets are created by the underlying 'exploitation of labour power' and are thus fictitious in nature. This conclusion suggests that financial products and services are always derived from the function of the real economy but are not always directly connected to real activity.

Also of interest is the focus on financial services within many developed economies, known as 'financialisaton', which can be described as 'the increasing role of financial markets, actors, and institutions in the operations of the domestic and international economies' (Fouskas and Gökay, 2012, p. 99). Part of this trend is the emergence of new forms of wealth creation, by financial means, perhaps using complex assets such as derivative instruments or credit default swaps (Fouskas and Gökay, 2012). This trend has been accompanied by an increase in the application of ICT, which has affected the proliferation of financial markets and wealth-creation strategies. It is unclear whether technological advancement led to financialization or whether the financialization trend led to technological advancement. This lack of clarity is summed up by Currie and Lagoarde-Segot (2017, p. 212). The area warrants further research, but this is complicated by the interdisciplinary nature of financialization (finance/economics, social sciences, and geopolitics). For the purposes of the present study, it is sufficient to note that the financialization and digitalisation trends have coexisted and progressively evolved since the 1980s. The direction of the causality is not relevant.

Along with the financialization trend, there have been changes to market structure through the increasing use of algorithmic trading and HFT. This phenomenon has been covered in previous chapters. It is worth summarising a few points made by O'Hara (2015). Firstly, exchanges now typically operate with their own computerised matching engine, which, in essence, is the exchange's order book. The order-matching process has a depth/price element, but where prices are tied, the longest sitting orders are prioritised, which places great emphasis on latency and has created new opportunities for latency-based strategies. Although HFT has likely had a profound effect on the strategies used and due to the latency characteristics of the markets, regulators seem to struggle to define HFT. For example, O'Hara (2014) argued that the SEC has only been able to define HFT based on a list of characteristics (as explained in Chapter 2). For example, in 2008, some NASDAQ traders were able to react to changes in the limit order book within three milliseconds (O'Hara, 2015). In addition, Chapter 2 has illustrated the differences in latency between traders, which can be problematic in terms of providing liquidity to slower traders, as illustrated by Stenfors and Susai (2019).

4.3 Fairness in financial markets?

The key question from the perspective of a market participant is that of fairness – or, at least, perceived fairness. Fairness is a difficult concept to define. Some may say that it is a question of a moral blemish, and others may say it is a matter of procedural fairness, while yet others may believe fairness to be a question of allocation of resources (Angel and McCabe, 2013). For this reason, the US SEC has no working definition of the term. The Dodd-Frank Wall Street Reform and Consumer Protection Act 2010 proposed a narrow definition of the term 'fairness', which only applies to the retailing of consumer financial products and is more akin to the concept of 'fair dealing'(Angel and McCabe, 2013).

Despite the difficulty in determining what might be unfair and the challenges of persuading others why this might be the case, some sources do go as far as claiming that markets that contain HFT activity are not fair to all participants.

Angel and McCabe (2013) highlighted that US network television has given a voice to people who claim that the ‘stock market is rigged’. Furthermore, Scott Patterson devotes almost all of the 335 pages of his book to illustrating the unfairness of modern trading environments, mostly using only anecdotal evidence. Patterson (2012) focuses on the increasing latency element, especially after the 2007 crash, which caused margins to become thinner. Some strategies, such as ‘spoofing’ (the submission of fake orders), were first detected around this time and are detailed in Patterson (2012). A firm named Trillium is known to have spoofed around this time with the aim of creating the illusion of other traders that a large buyer or seller was present in the market (Patterson, 2012). This is an example of a vexatious momentum ignition strategy. Along with thinning margins, US markets were also less volatile in the post-crash period, which led to the spread of technology into other financial centres, particularly in Asia. In 2009, the National Stock Exchange of India began provisioning for co-location facilities to a move to court this type of activity. Thus, the incentive to engage in aggressive behaviours and vexatious strategies was driven by a search for yield on the part of traders and trading venues.

The research of Harris (2013) illustrates that negative externalities can be generated in markets characterised by HFT activity. For example, front-running strategies are often considered to ‘hurt’ other investors. This must be taken with the increasing drive to reduce latencies that make asymmetries wider and potentially make negative externalities more severe.

In terms of HFT activity, Harris (2013) distinguishes between valuable, harmful, and very harmful activity. On the one hand, HFT activity provides liquidity to markets and has reduced transaction costs for many investors, although it must be noted that liquidity can be short lived (Stenfors and Susai, 2019). On the other hand, harmful activities include attempts to incorporate information into price before others may have chance to update their own limit orders, which is known as toxic arbitrage (Harris, 2013; Foucault et al., 2017). An example of very harmful activity given by Harris (2013) is the activity of anticipating order flow from other traders and trading against this anticipated flow.

Thus, there is an argument that slower traders would welcome restrictions on faster traders. However, many of the vexatious strategies HF traders use are legal, and the option to use these strategies is open to any market participant who is able to use them. This counterargument gives rise to the suggestion that a rule that is applied to all in a market may possibly be seen as a fair rule. This is the basis of ordoliberal regulation, which may be helpful in exploring how measures against HFT could be designed if needed.

4.3.1 Ordoliberal regulatory approaches

The ordoliberal view of markets stands in contrast to neo-liberal thought, which became popular in the 1960s and was further championed in the 1980s by politicians such as Margret Thatcher and Ronald Reagan. A related ‘variety of capitalism’, the German–Austrian perspective of ordoliberalism, has also been labelled as ‘liberal authoritarianism’ (Fouskas and Roy-Mukherjee, 2019).

The crux of applying ordoliberal principles to financial regulation is the concept of the state’s ability and willingness to impose conform interventions on markets. These interventions do not alter a market’s tendencies for profits to diminish in the long term or the ability for firms to increase profits in the short term (by increasing production or cost cutting). The core of the idea is the somewhat paradoxical concept that markets with more regulations may in fact be freer markets.

Conform interventions, according to Siems and Schnyder (2014), are a mixture of regulatory actions and ordinating actions. Regulatory actions are aligned to market microstructures (sources typically refer to the elimination of monopolies) and the promotion of price stability. In the HFT context, this could mean using measures such as minimum order resting times or requiring exchanges to operate circuit breakers in order to deter the use of strategies that cause volatility or to limit the extent of extreme price movements.

Ordinating actions, on the other hand, are the preconditions for market structure. This is not a total break away from a laissez-faire ideology but rather the separation of laissez faire into two strands. The first is the separation of markets and governments, which is broken, and the second is freedom to participate in a market, which is still held (Siems and Schnyder, 2014). This split in freedom in general does not sit hand in hand with the concept of laissez faire, which focuses on minimum government intervention (or interference). Indeed, some who subscribe to the ordoliberal view believe in the concept of the ‘strong state’, which is able to provide conform-type regulations. Overall, it is not possible to argue that a system is either and only ordoliberal or laissez faire. Rather, these doctrines offer theoretical bases against which market interventions may be considered.

For example, the ordoliberal philosophy also emphasises minimising an individual’s market power, possibly minimising the ‘too-big-to-fail’ concept (Siems and Schnyder (2014). This is a moot point, because the non-collusive oligopolistic structure seen in some financial sectors is often organic in nature and is largely held due to regulatory requirements and where minimum capital requirements that are very difficult for new entrants to meet apply.[17] Research in this area applies to retail banking; however, this is not directly relevant to the present study, as retail banking requires a great degree of consumer protection and poses rather different regulatory challenges. Cruickshank (2000) and Cable (2014) are interesting sources for those interested in the retail banking sector.

A final point to note on the forms of intervention drawn out from the ordoliberal perspective is the variance of adoption between countries (Siems and Schnyder, 2014). This gives rise to a phenomenon known as regulatory arbitrage. In principle, when an asset is traded in multiple locations, this can be a relevant issue. This is a developed concept in the field of banking regulation, whereby banks seeking light-touch regulations are able to transfer funds to markets with weaker regulations or undertake off-balance sheet activity (Houston et al., 2012; D’Avino, 2017). This concept is illustrated in the over-the-counter derivatives markets for swaps and futures between the US and EU member states in response to tightening regulations in the US (D’Avino, 2017). Thus, the issue is present in financial markets; however, evidence from exchange-intermediated markets is not available. Data from Bloomberg’s equity-screening function illustrates that there are 204,504 equity securities listed on Western European exchanges, which are cross-listed with at least one other equity listing elsewhere in the world. This idea illustrates that the scope of dual listing and dual authority is not minor. For instance, where shares are cross-listed, price volatility is arbitraged into two marketplaces and may be more persistent if measures taken against

[17] Barriers to entry could be imposed by the government or regulator. Trading venue rules may also impose restrictions like listing requirements.

this arbitrage are asymmetric. This is an area that requires further research and is beyond the scope of this thesis. The issue that is directly relevant here is the idea that market abuse can have global elements, and consistent global regulation would be one potential solution to that problem.

4.3.2 Global regulation

The question of global regulation in the ordoliberal view is not well developed, particularly as the ideology dates back to the 1920s and 1930s (Bonefeld, 2012; Siems and Schnyder, 2014). There is an argument that the idea of the strong nation state is not compatible with external/transnational regulation. However, it is noteworthy that much of the ordoliberal literature does not consider the cross-border nature of an intangible good such as a financial instrument.

Thus, two questions arise:

1. Should global or regional regulations in financial markets be considered?
2. What form would these regulations take if they existed?

Kerwer (2005) gives an example to illustrate the principal difficulty in applying regulation across international borders: the ban of the International Air Transport Association against smoking on passenger flights. Aeroflot, a large Russian airline, was reticent about this requirement, as smoking cigarettes is culturally normal in Russia, and preventing passengers from smoking would be a restrictive practice to Aeroflot. This example shows that the fairness or necessity of a regulation is related to localised considerations about what constitutes adverse spill over or negative externality.

As in this Aeroflot example, global financial regulation is not law; rather, it is a global standard, which is not a legal requirement unless a national body opts to enact a standard in the form of national law. At the core of setting standards is the aim of convincing rather than coercing firms to adopt best-practice principles (Kerwer, 2005, p. 611). Financial standards seek to encourage firms to control the risk arising from their activities internally. Examples do exist of super-national regulation; however, their success can be tempered by differing regulatory traditions and by the concept of national sovereignty (Kerwer, 2005). Regulatory structures and the unique case of the EU will be discussed later in this chapter.

Also of relevance is the concept of governance by standardisation. Again, standards may be legally adopted, but they are not legally binding in and of themselves. Setting standards overcomes the need to implement a device that is legally binding in multiple countries and has become, according to Krewer (2005), the basis of the coordination of global policy following the 2007 financial crisis in the form of best-practice rules that are applied by national regulators. Leaving implementation to national regulators builds in some element of flexibility and allows each regulator to implement its own systems concerning monitoring and enforcement practices. Examples include the International Monetary Fund’s (IMF) multilateral surveillance of national policy, which may disrupt exchange rate stability. In addition, the remit of the Financial Stability Forum, which was replaced in 2009 by the Financial Stability Board, is to make recommendations about the wider global financial system. This is housed within the Bank of International Settlements (BIS). The Financial Stability Board has an interest in systemic risk and stability and illustrates that some level of global cooperation is possible – although not to the extent that any one jurisdiction can be compelled to action.

Also within the BIS’s remit is the development of the Basel Accords. The Basel Committee on Banking Supervision was established in 1974 following the collapse of the Bretton Woods System (1971), a stock market crash in 1973, and the oil price shock of 1974. Twenty-seven countries are members of the committee; however, other countries have chosen to adopt substantive elements of the Accords. The Basel Accords on capital adequacy have evolved over time. Basel I was issued in 1988, Basel II in 2004, and the latest amendments led to Based III in 2010 (full compliance with Basel III was required by January 2019) (History of the Basel Committee, 2021). All three sets of accords revolve around three pillars/principles: supervision, market discipline, and capital adequacy (Regulation Guide: An Introduction, 2011).

Firms must address each of the three elements in the following ways:

1. National governments and/or regulators must set the minimum capital requirements; ergo, firms must comply with these requirements by meeting or exceeding local requirements. It is important to note that this element creates some scope for regulatory arbitrage though subsidiary activity.

2. Concerning the supervision of individual banks, individual regulators must create and implement a risk-management process based on these minimum requirements in order to set minimum capital requirements that are commensurate with an individual bank's risk profile.
3. Finally, in terms of market discipline, individual banks should be held to standards of maintaining transparency through disclosure to stakeholders (including shareholders, clients, and researchers).

Another example of international coordination is the Global Investment Performance Standards. Compliance with these standards is voluntary. The standards set out methods for the calculation of the performance of investments. The key objective of the standards is to make investment prospects comparable regardless of multiple locations, as the idea is to create a single global standard. The goals are accurate calculation, fair disclosure, and global standardisation. A limitation of these standards is the need for firms to voluntarily adopt these principles. At the time of writing, the author is not aware of a country that legislated these requirements. Compliance in this example is self-certified, although firms may be encouraged to have their compliance validated by an external audit.

What becomes clear is that most of the global standardisation efforts are optional and unenforceable, as they are not reflected in national laws, with the exception of the Basel Accords and Directives made within the EU. Ergo, the scope of international efforts in this regard is mixed.

4.4 High-frequency trading control devices

The relevant literature contains many proposed or implemented measures to curtail the activities of HF traders in various ways, which can broadly be categorised into three types: exchange architecture devices, price/payoff interventions, and regulatory intervention/contact. All these interventions are theoretically within the remit of the regulator to impose, although many exchanges can voluntarily adopt these control devices using their own discretion and guidelines, because trading venues control so much of the market microstructure for assets that are not traded over the counter. This also means that markets such as ForEx are harder to regulate, as few control devices are available.

4.4.1 Exchange architecture

Exchange architecture refers to the rules and practices in use in trading venues. These restrictions are most relevant to exchange-intermediated markets (such as equity markets); however, some elements can be replicated in trading systems for over-the-counter markets such as the ForEx market. Examples from the literature include: circuit breakers, limits on order submissions, minimum order resting times, restrictions on short selling, and 'speed bumps' in trading. These devices are designed to alter the incentives of traders through non-direct monetary means. Most of these devices aim to reduce the extent to which a latency advantage can be beneficial to a trader. Circuit breakers are an exception to this principle, as they are designed to react to changing market conditions and restrict market activity accordingly.

A circuit breaker is a mechanism designed to halt trading upon the detection of the occurrence of a prescribed event. The duration of the halt brought about by the circuit-breaker opening varies, as does the pre-condition to open the circuit breaker. The core idea of the circuit breaker is to cause trading activity to stop, which gives traders time to gain a broader understanding of the causes of extreme price changes. This mechanism is not designed to restrict the activity of HF traders; however, it does warrant discussion, as HFT activity is known to exacerbate extreme volatility events. As a result, a mechanism to stall an extreme volatility event (Chung and Lee, 2016) does account for the nature of traders and may have increasing value as lower latencies are seen in trading venues. This intervention is also capable of preventing an 'order imbalance' that may be brought about by extreme downward price pressure caused by a stochastic event or an erroneous order submission (known in the trade as a 'fat finger' error) (Subrahmanyam, 2013).

For example, when faced with rapidly falling prices (i.e. a decrease equal to 5% of the opening price), a large amount of short-selling activity is likely to occur, which, on many exchanges, will trigger a short (perhaps 90-second) opening of a circuit breaker. However, the halt of just 90 seconds serves two purposes: Firstly, it sharply brings the price movement to traders' attention, and secondly, the halt provides time for all parties to update their information set and perhaps more fully understand what triggered the downward price pressure.

The conditions for a circuit breaker to open and the length of the halt vary across exchanges. In order to illustrate the typical parameters, an example of the parameters on NYSE is used. NYSE, as of 2013 (rule 80B) holds that, if the S&P500 index or the DJIA index values fall below a certain threshold (typically an absolute price movement of 10%), then a halt takes place. In the case of NYSE, a tiered structure exists. In the first instance of a 10% drop in index value, a halt is applied, with the duration determined by the time of day. Before 14:00 the halt holds for an hour. Between 14:00 and 14:30, the halt is 30 minutes. After 14:30, no level-one halt is applied (Subrahmanyam, 2013). For a 20% decline in index value, a second halt is triggered by the opening of a level-two circuit breaker, whereby a trading halt is initiated for two hours when the circuit breaker is triggered before 13:00, and a one-hour halt is triggered between 13:00 and 14:00. After 14:00, the level-two circuit breaker ends trading for the remainder of the day. In the event of a price decline of 30% (the third circuit breaker to open), regardless of the time of day, trading is halted for the remainder of the day (Subrahmanyam, 2013).

In January 2016, the two principal Chinese equity markets, the Shanghai Stock Exchange and the Shenzhen Stock Exchange introduced a system of circuit breakers applied to the CSI 300 index. In this instance, two circuit-breaker levels exist at price movements (±) of 5% and 7% respectively. In addition, an individual asset may stop trading after a price movement of 10%. The level-one circuit breaker halts trading for 15 minutes unless it is activated in the latter portion of the trading day, whereby the halt lasts for the remainder of the day. The level-two circuit breaker suspends trading for the remainder of the day, regardless of what time the movement occurs (Wang et al., 2019).

There is evidence that the opening of a circuit breaker does not, in practice, halt price declines; however, it is likely to provoke trading activity and raise volume traded. In addition, it has been suggested that the circuit breaker does not impact the bid-ask spread (Wang et al., 2019). This conclusion is based on an assessment of data on the two above Chinese exchanges between 04 and 07 January 2016, during a market volatility event. Caution must be taken here: This is a study of a single connected event that did show persistence. Taking this finding and applying it to all cases of circuit breakers would not work.[18]

There is also a related phenomenon that causes the existence of a circuit breaker with clearly defined trigger points, known as magnet effects. The prices may accelerate towards the bounds that trigger the circuit breaker; however, prices slow (then hold or reverse) before triggering the circuit breaker (Cho et al., 2003; Wang et al, 2019). This is likely due to the universal knowledge of the circuit breaker and the general desire of the majority of traders not to provoke a breaker to open. Cho et al. (2003) found that the phenomenon is more visible at the upper bound (prices increasing); however, a weak relationship with the lower bound (prices decreasing) is also shown. The circuit breaker thus may provide price stabilisation by virtue of its existence rather than its operation, which is largely a positive concept, as it may be desirable to regulate price without triggering a trading halt. On the other hand, it is necessary that prices reflect as much information as is practical,[19] and it is not clear what effect this generalised hesitancy to provoke a circuit breaker has on price being a true vector of all information.

A retrospective analysis of the data from the 1987 market crash illustrates the role of a circuit breaker in increasing the information basis on which trades are made. The wider context is markets that adopt a continuous trading model when decisions are taken based on a small information set i.e. a poor understanding of why the current price is the current price (Greenwald and Stein, 1991).

4.4.2 Limits on order submissions and cancelations

It is also possible to consider devices that are non-reactive in nature. For instance, the use of an OTR provides an alternative method that sets targets for firms regarding executing orders rather than cancelling them. In these cases, there can be a financial penalty when an OTR target is not met.

On the one hand, it is known that HF traders tend to have a very high OTR due to their rapid order submission and updates, and limiting this practice could curb the practice of order submission and withdrawal. This limitation would in turn reduce the level of

---

[18] The May 2010 flash crash provides a contrasting example, which is discussed later in this chapter.
[19] This is perhaps a rather pragmatic view of the efficient market hypothesis.

noise in financial markets. As shown in Chapter 2 of this thesis, noise does exist in markets but not at a level that is likely to make market participants' trading strategies trembling-hand unstable. Ergo, while a policy that aims to reduce noise in markets would be effective, the necessity of doing so is not total.

Nevertheless, a potential benefit is the barrier to some of the known vexatious practices within HFT. Activities such as spoofing and quote stuffing involve the placing and withdrawal of orders with the express intent that the orders will never be executed. The aim is to manipulate the price or slow the speed of the exchange's matching engine. A consequence of both activities is an increase in the trader's OTR. Hence, an OTR limit would at least make algorithmic traders use momentum ignition strategies more sparingly.

An example of OTR restrictions can be found concerning Borsa Italia, the principle Italian stock exchange (based in Milan). The Borsa charges a fee to all traders who have an OTR in excess of 100:1 of €0.01 per order, and for those with higher OTRs the fees increase to €0.02 for an OTR above 500:1 and €0.025 for an OTR above 1000:1. Fees are charged daily, up to a maximum of €1,000 per trader (Chung and Lee, 2016). Further examples using the Nasdaq methodology are available (Nasdaq, 2019).

As of May 2016, the rules of the EUREX exchanges in Germany and Switzerland (Zurich) have been amended to require the calculation of a trader's OTR on a monthly basis, and the OTR is analysed on the last trading day of the month (Peters, 2016).

Other examples exist within Europe. For instance, France levies a tax of 0.01% of the value of cancelled orders. In addition, Norway has introduced an OTR limit on the Oslo Børs, which, similar to the Italian system, places a charge per order of 0.05 Norwegian krone on trades with an OTR in excess of 70:1, and this is billed monthly. The system in Oslo seeks to only penalise certain practices, and, as a result, liquidity-providing orders are not counted in the OTR. Certain order types, such as execute and eliminate or fill or kill orders are affected by the system in place in Oslo; therefore, there is a financial disincentive against using these order types. In addition, an order that remains in the market for more than one second is not counted in the trader's OTR. It is clear that the specific aim of these rules is to minimise practices used by HF traders, some of which can be vexatious yet legal, although some HFT activity is, of course, harmless.

As mentioned above, on the Oslo Børs, an order that remains on the market for one second or more is not seen as market abuse and is not counted in a trader's OTR. This rule leads the discussion to the more general use of minimum resting times for orders, which appears to aim to remove incentives to update orders rapidly. The rule would likely would have a positive effect in reducing toxic arbitrage[20] (Foucault et al., 2017). This market intervention would also likely prevent strategies such as spoofing, as the minimum period increases the likelihood that an order would be matched unless it is placed deeply in the order book. HF traders are known to withdraw orders as the market updates, and, as long as they withdraw orders due to price and information updates, the practice is legitimate. Such a policy would also likely restrict the legitimate activity of HF traders, including those that are tasked with liquidity provision in the role of the designated market maker.[21]

At the time of writing, the author is not aware of an example of this idea that is actually in operation. Chung and Lee (2016) mention the Nasdaq OMX PSX exchange once offered minimum-life orders, which required a minimum resting time of 100 milliseconds. As an incentive to use this order type with a minimum resting period, a higher liquidity rebate was offered within the maker-taker structure – in this case, 0.0026% rather than 0.0024% per share. This order type is no longer available on the exchange.

A slightly different way of slowing down activity is to slow all traders' order submissions to the exchange. Various exchanges have taken a slightly different approach to add a lag to all trades, simply to add latency between the submission of the order and the potential execution (Lewis, 2015), dubbed as speed bumps in trading. In 2017, the NYSE gained SEC approval for a policy that adds a delay of 350 microseconds to orders against small-to-medium capitalised companies' stock.

Similar practices are known to be in use on the following exchanges: Par FX, EBS Market, IEX, Aequitas Neon, TSX Alpha, Refintiv Matching, Eurex, NYSE

[20] Defined as trading against quotes that have yet to update to new information.
[21] For a description of this role, refer to Stenfors and Susai (2019) or the discussion in Chapter 2.

(American) Nasdaq, Inelegant Cross, and Moscow Exchange. In this list are venues dealing in ForEx, equity, and futures (Osipovich, 2019).

This practice, as is clear, is in rather wide use; however, it is neglected in the literature and is perhaps worthy of further examination. Considering the concept of the importance of the first mover's advantage, it is possible that traders are working fast to gain the first mover's advantage and obtain a favourable position in the queue to execution. Ergo, a speed bump in trading is indeed a practice that could slow traders down, but only in relative terms, as the race to the front of the queue would still remain.

For example, in a group of runners in a 100-metre sprint, each runner runs as fast as possible, and they are ranked based on their race time. If an arbitrary amount is added to each runner's time, perhaps one second, this has no effect on the order that the runners crossed the finish line. If all runners are handicapped in the same way, then Usain Bolt would still hold the world record; however, he would be one second slower (9.58 vs. 10.58 seconds). This is a metaphor for the race to the matching engine in which HF traders engage when new information becomes available. The only way to address this issue would be to apply restrictions only to faster traders; however, a decision concerning which traders require lagging would be arbitrary and, according to many views, 'unfair'. Consequently, this policy would likely not be effective. Furthermore, it would not be compatible with either a laissez-faire approach or an ordoliberal approach to market regulation.

A strategy that is observed in practice is the imposition of restrictions on short selling. These restrictions essentially make the practice of short selling, or at least certain short-selling strategies, unavailable. It may be possible for certain exchanges to adopt restrictions only upon naked short positions, dependant upon what position checks they make upon members. In some cases, these restrictions have existed only for short periods of time in response to marketplace conditions.

In the UK, these circumstances are laid out in the short selling regulations (Short Selling, 2020a). These were originally derived from an EU Directive, however their current basis is likely the Short Selling (Amendment) (EU Exit) Regulations (UK, House of Commons, 2018). In brief, these regulations give the relevant regulator authority (currently the FCA) a mandate to implement regulations when either there is a significant price fall or 'exceptional market conditions' are observed. A price decline is measured as the percentage decrease based on the previous day's closing value. It is not clear what the percentage threshold is. It is also not clear if short selling restrictions are or have ever been used in a coordinated way alongside circuit breakers, as they are both interventions triggered by significant price movements in an asset or an index. However, their would be a logic in doing so, as this would prevent HF traders from deploying the strategy en masse in market conditions which are unusual and where slower situational awareness may be lacking.

In addition, short selling restrictions in the UK can be applied in cases of 'exceptional market conditions'. These are defined as 'adverse events or developments that pose a serious threat to financial stability or market confidence' (Short Selling, 2020a). The May 2020 flash crash event (outlined on page 120), is a good example of such a scenario.

The short selling regulations in the UK context derive from an EU Directive; thus, similar powers are available to the regulatory authorities in all other EU member states. Although these are EU-derived regulations, there is no indication that they will cease to be applied in the UK beyond 2021 due to the provisions set out in the European Union Withdrawal Bill[22] (EU [Withdrawal Agreement] Bill, UK, House of Commons, 2018) that address the transfer of EU law into UK national law (Lang et al., 2017).

The powers of interventions are the same in all EU member states. The only item the FCA guidance notes add is that the duration period of an intervention is at the discretion of the appropriate regulator in each country (Short Selling, 2020b). No notes or guidance are provided to indicate that regulators may take a harmonised approach, which may be helpful, as assets are traded in multiple EU member states.

It is interesting that short selling restrictions could be used in connection with circuit breakers to restrict downwards price pressure without the need to halt trading. No evidence can be found to illustrate this is in effect.

In theory, it may be possible to apply a short selling restriction as part of a trading venue's system of circuit breakers. For example, consider a circuit breaker with a

---

[22] Some media sources refer to this as the 'Great Repeal Bill', but this is not the correct name.

device to place a short-term halt on short selling. This would have the effect of limiting the ability to deploy downwards scalping strategies and volume-driven strategies while allowing information to still be traded into price. This may act as an initial intervention that could be followed by a full trading halt at a later stage if necessary. By combining these policies, making circuit breakers more sensitive (as necessary) to control market volatility events would be justifiable, as the opening of a circuit breaker would not immediately lead to a halt in trading. It is important to note that a halt in trading is likely undesirable as it may impose indirect costs on traders and undermine confidence in the affected trading venues.

Further evidence in this area can be found in evidence from practices in the US. In the US, short selling was made possible by the SEC's regulation 'SHO' in 2005 (Regulation SHO, US Securities and Exchange Commission, 2005). Beforehand, US markets were subject to the uptick rule, which prevented the submission of orders for short positions during periods of downwards price pressure. This rule had been in place since the 1930s.

The decision to allow for wider practice in terms of the use of short positions was based on a permission for a trial period – with a period of observation granted by the SEC – to allow for the study of a permanent removal of restrictions. It was found that the facility to short sell, even on 'down ticks', allows for better incorporation of information into price where an event is well known in the future, but there is also the risk of 'coordinated short-selling attacks' (Grullon et al., 2015). Grullon et al. (2015) evidence that the impact of short selling placing downwards pressure on equity prices can lead to constrained investment decisions on the part of the issuing firms in equity markets (Grullon et al., 2015). This finding illustrates a spillover (negative externality) effect directly arising from this practice; in fact, this is one of the clearest cases in which a trading strategy can be shown to be vexatious.

4.4.3 Measures for controlling prices and payoffs

In addition to the available measures of altering behaviour by altering the exchange architecture, there are strategies that affect the price of an asset more directly. There are two general forms of intervention: transaction taxes and rebate structures.

A financial transaction tax could be applied on a per trade basis, perhaps targeting certain activities, adding a cost penalty to a trader for each trade they make. The logic of a tax per trade may particularly affect HF traders who, by nature of their typical strategies (such as scalping), submit many orders though the course of the day. In addition, the literature reviewed in Chapter 2 shows that the margins per trade for an average HF trader are both thin and have been declining over time. Thus, the addition to the cost per trade reduces profit per trade, and this may require HF traders to avoid certain activities altogether. It is notable that the method by which the tax is applied matters; for example, taxing a percentage of per-trade profit (i.e. 5%) would have little effect on HF traders – this approach would be consistent with the principles of rules-based capitalism. On the other hand, a tax of a fixed amount per trade (for instance, $0.001) would be extremely restrictive to HF traders and would necessitate either a review of their HF strategy or the use of latency-based technology in general.

As a tax per trade would increase the overall trading costs, the depth of the margin per trade would need to increase in order to make trades profitable. It is likely that, if traders needed to find thicker margins, then they would trade less frequently and desist from scalping strategies and some market-making activity. HF activity tends to tighten the bid-ask spread (see Chapter 3; Narayan et al., 2015). Techniques such as scalping would remain possible but would require greater time periods, as the spread capture would require a greater number of ticks. In general, this approach may slow down HFT activity by diminishing the scope for latency arbitrage.

The tax revenue generated from this form of taxation is generally made available to central government, and Chung and Lee (2016) argued that this revenue can be used by governments to partially offset their obligation to act as the lender of last resort during a financial crisis that threatens the stability of the overall financial system. The question arises concerning to what extent this tax could be considered Pigouvian – the question that may decide the effectiveness of this strategy – however, it provokes a discussion concerning the positive and negative externalities of trading financial assets via exchanges.

The alternative option in terms of pricing interventions is to consider the rules within the trading venue. Many exchanges use a trading fee rebate structure to incentivise liquidity provision; that is a rebate to the side which matches a standing order. This approach must, however, be tempered by the incentives it gives to HF traders.

Another option is an unequal split in the fees charged rather than the use of rebates. Both methods are practical and are seen in practice. For example, in a scenario where the exchange must incentivise a bid to sell, consider two scenarios:

1. The exchange charges fees of 0.006% upon asks and 0.004% upon bid orders.
2. The exchange charges fees of 0.015% upon asks and makes a rebate of (pays) 0.005% upon bids.

It is important to note that these examples both have a net commission to the exchange of 0.01% but achieve this in two different ways. At prima facie value, the second option, which offers a rebate to those who provide liquidity, may be more effective, as it is a direct incentive to provide liquidity.

However, the efficient market hypothesis conjectures that all known and available information is incorporated into prices by way of the trading process. Thus, it would be possible that prices in the exchange reflect the fees and rebates that are available and that traders in these markets make decisions based on the effective prices that they know they will pay in clearing rather than the prices listed in the exchange. Maker vs Taker Fees (2020) supports this conjecture by presenting an example that essentially shows that Knowledge of the fee structure leads to a widening or tightening of the order book's display, as market participants adjust their bid and ask prices accordingly. It is noteworthy, however, that there is a caveat to this premise: High fees can prevent an effective tightening of the bid-ask spread. For instance, if there are taker fees amounting to 0.15%, the actual spread would need to be 0.3% or greater for this form of mental accounting to be effective. This conclusion is based on the principle that the actual bid-ask spread must always be strictly positive (markets lock when the spread is zero). As markets are generally more efficient with a tighter bid-ask spread, this approach of taxing transactions may create one problem in exchange for solving another.

4.5 Regulatory contact

In the case of the UK, it is within the remit of the FCA to supervise HFT activity and initiate action against those who commit acts of market abuse. An FCA report states, 'We continue to proactively supervise algorithmic trading activity and conduct research on algorithmic trading' (FCA, 2018, p. 3). Based on an internal report, cited above, the supervision approach is specific to each firm and contains an element of awareness training, with the expectation that firms can add the prevention of algorithmic market abuse to their internal compliance remit.

In UK markets, exact standards are set out in chapter 7A of the FCA's Market Conduct Handbook, and these standards are what the FCA seeks to enforce (FCA, 2018). This handbook gives a detailed account of the FCA's requirements of a firm's internal monitoring. Although it does suggest that the FCA actively monitors firms, it does not specify the extent thereof, and this point remains unclear.

Examples of misconduct involving HF traders are not numerous, yet they are not difficult to find. In many cases, evidence of misconduct and market manipulation is found through the process of monitoring. One such example is the US regulator's prosecution of Deutsche Bank, HSBC, and UBS for spoofing. In addition, US authorities arrested and charged eight individuals who were instrumental in the execution of the illegal strategy.[23] The extent to which active monitoring detected this activity is unclear, as it is known that some of the evidence used was text messages between the individuals, and it is noted that UBS self-reported the extent of the activity (Jopson et al., 2018). Monitoring may have instigated a wider investigation.

The limitations of the monitoring capability have been further exposed by the FCA's admission that it has been unable to find evidence of strategies such as front running in UK markets (Stafford, 2016). Furthermore, the FCA cannot distinguish between fast reaction times and the ability to predict order flow when it is only analysing tick data (Stafford, 2016). This position supports the view that regulators need better access to data, perhaps in the form of a 'consolidated tape', which, as a concept, has been

[23] In the USA this is an offence as set out in the Dodd Frank Act (2010).

available to regulators in the US since 1976 to provide summary bid-ask quotes across exchanges in the US (Consolidated Tape Association, 2020).

Access to data is likely to be a key issue in the future of financial regulation, as the ability of regulators to use this data is likely to increase. Furthermore, it is likely that regulators may use data in a proactive way, i.e. in near real time, rather than in a reactive way in order to identify potential market abuse using a technology-based approach. This approach is known as RegTech.

The term RegTech refers to the use of technology to assist the regulation of markets and the monitoring of compliance with regulations and market rules. The concept has existed as long as electronic markets have existed (Nasdaq was the first in 1971) and has been of interest as long as algorithmic trading has been known to lead to occasional market volatility events. The first significant event is thought to have occurred in 1984 according to Arner et al. (2016). Of course, interest was once again renewed following the financial crisis of 2008 and the flash crash event in May 2010. Another factor that has influenced interest in RegTech is the adoption of frameworks such as the Investment Services Directive in the US and the Markets in Financial Instruments Directive (MiFID), which both emphasise the role of data and the separation of data and transparency (Arner et al., 2016).

Both fintech and techfin[24] are terms that reflect the technology applied in financial firms and technology firms venturing into finance. As firms in financial services are developing a much more collectivised approach to data technology, which can be used for internal compliance (a function of individual firms) but can also be used in an external monitoring role, subject to the proprietary role of data.

Arner et al. (2016) reported that the FCA believes that RegTech is a subset of fintech; however, they argue that this is a limited view of what RegTech is. While RegTech is viable in terms of internal compliance and monitoring, it can also monitor submissions to regulators (using AI and machine learning technology). In addition, the use of technology offers regulators the scope to increase the continuity of monitoring with favourably scalable costs or economies of scale.

[24] There has been rather rapid growth in the use of this term in China.

This line of enquiry is taken further by Keller (2012), who argues that RegTech has the scope to increase market transparency by more actively scrutinising markets and visibly dispelling assumptions that markets are just ‘a bunch of computers which are out of control’ (Keller, 2012). This approach could also be part of a wider move to pressure traders to disclose, in general terms, what activity they are conducting. In the US, SEC rule 13h-1 requires certain traders (including most HF traders) to make some disclosure of their activity. More active monitoring is hoped to discourage abusive practices and to allow greater data capture and analysis to be used in retrospect of incidents such as flash crashes or when allegations of misconduct or microstructural irregularities have been alleged.

4.6 The May 2010 flash crash incident

It may be helpful to examine one of the most notable recent cases of a market volatility event. As this is a matter of pricing, it may be wise to hesitate before calling this a market failure, as to do so would raise the question of the exact meaning of the term ‘market failure’.

On 06 May 2010, a market volatility event took place, commonly referred to as the flash crash. Whilst the term ‘flash crash’ often refers to this specific event, the term can also be used to describe a market volatility event (a sudden price movement that, for most traders, is unexplained).

In this particular flash crash, a market volatility event took place within the market for S&P 500 e-mini futures. Over the course of half an hour, the index futures experienced a 5% price decline followed by a prompt reversal in price momentum. The incident began with a large put order of 75,000 units. Two points here are relevant: Firstly, such a large order, with reasonable certainty, would make a price impact, and secondly, the order is very unlikely to have originated in an algorithmic trader, as such traders are designed to shred large orders. This appears to have been an error made by a human being. This form of error is commonly known as a ‘fat finger’ error in the profession. The result is, of course, an order book imbalance and a decrease in asset price (Easley et al., 2011).

This situation subsequently leads to a liquidity imbalance at both the index level and on an individual level for some stocks. Easley et al. (2011) postulated that there are three microstructural elements that are significant in this matter:

1. Liquidity provision is concentrated to a small number of traders. In many markets, these traders are HF traders who take on the role of designated market maker. An associated factor is the inventory-neutral nature of these traders, who hold low inventories. This is especially the case when systems are designed not to hold inventory overnight and remain close to inventory neutral throughout the day.
2. The share of HFT activity, along with the decreasing active participation of institutional investors, is relevant. This parameter refers to the proportion of the market that follows scalping strategies and the relative proportion of the market that follows a more typical capital-gain approach[25] or even an income approach.
3. High sensitivity to daily profit and loss combined with the thin margins in HFT leads to the use of aggressive algorithms. Easley et al. (2011) did not develop this point. It is the author's opinion that aggressive algorithms were being used in a way that did not account for the wider market conditions and had little scope to react to information other than price and predicted order flow.

During the course of this event, it is believed that 16 trading accounts were participating using HFT technology, and together, those accounts generated a third of that day's trading volume. In their retrospective analysis of the event, Kirilenko et al. (2017) found that the HFT traders did not modify their behaviours or strategies in response to the price decline and subsequent market volatility. Instead, they used an aggressive strategy of taking liquidity and short selling. This created a scenario whereby the HF traders were short selling, and slower fundamentalist traders were taking long positions.[26] Again, using retrospective analysis, Kirilenko et al. (2017) argued that the downwards price pressure was caused by the aggressive nature of the HF traders: the volume at which they can operate and the possible ability of an HF trader to anticipate prices. The article mentions anticipating prices, although it is more

[25] The core distinction between scalping and capital gain is the time horizon and likely the willingness to use short-selling strategies.
[26] It could be assumed they were following a typical capital gain approach.

likely that an HF trader can anticipate order flow and deduce price from this information.

A further complication is the nature of this particular exchange, which trades 24 hours a day with only occasional technical halts (Kirilenko et al., 2017). While this fact has not been examined in depth in the relevant literature, it raises an interesting question that may warrant further research. In markets with a close period, it is thought that most information is incorporated into price in this period (Barclay and Hendershott, 2003). This overnight close possibly acts as a reset period in which traders of all types can update their information set without the latent effects of doing so being a consideration in terms of lost trading activity. In a hypothetical case of low-level market volatility that is short of a circuit-breaker intervention, this volatility could, however, continue for more than a trading day.

In the case of the May 2010 flash crash, approximately 30 minutes after the diminishment of cross-arbitrage and around 1 minute after the beginning of the steepest price drop, a circuit breaker opened and halted trading in the venue. In the period before the circuit breaker opened, assets that had been traded in other venues and on the affected exchange were exhibiting ‘impaired arbitrage’, especially in the half hour before the opening of the circuit breaker.

The May 2010 flash crash is not a unique event. It happens to be very well documented in reliable sources and therefore is easy to discuss. To illustrate that this type of volatility event need not apply to an index but also to a single asset, the Jardine Matheson volatility event could be considered.

On 24 January 2019, shares in Jardine Matherson began trading at prices that were far below normal on the Singapore Exchange, trading at US$10.99 given the previous day’s closing price of US$66.47. Around this time, Goldman Sachs Group Inc and Morgan Stanley were known to have placed large put orders, but the causation cannot be proven. At this point, many call (buy) orders were made, and this created an order book imbalance. The trading venue chose not to cancel the orders; rather, the situation resolved when agreements were made to settle the trades for a higher strike price (The Standard [Hong Kong], 2019).

4.7 Additional notes and lessons from Chinese markets

A small number of English-language sources illustrate that there are some differences between HFT practices in Western countries and in China. Many of these differences are related to market structure and the practices of the trading venues. These practices likely affect trading behaviour and thus may lead to different outcomes in terms of externalities and vexatious strategies. It is worth noting that most HFT activity in China takes place in futures markets.

The first key difference is the mandate for a position check, which is standard practice in China. This check has two elements. Firstly, a check is undertaken to confirm that a trader's order will not cross another order that has been submitted by that trader. Secondly, a check that the trader's futures verified account (end-of-day profit and loss is added/taken from here) is in credit is made prior to execution (Wang and Zheng, 2015). These position checks are not required in Western markets and do not seem to be conducted on a voluntary basis at exchange level. Indeed, it is thought that many HFT systems do not make an internal position check to reduce computational burden and decrease latency.

In addition, it is important to note that transaction costs are higher in Chinese markets – a typical value is around 0.0025% of the traded asset's strike value (Wang and Zheng, 2015). This likely makes the scalping strategy less viable. Although transaction costs may be higher, China does not have a system of capital gains tax as in many Western countries, and this may incentivise trading activity by way of reducing costs (Wang et al., 2019).

Finally, market data feeds are rather different in terms of the frequency at which they update. In Western markets, data feeds update based on order flow; in many cases, this practice results in updates on a millisecond-by-millisecond basis. In the Chinese futures markets, updates are only provided every half second (twice in one second) (Wang and Zheng, 2015). Wang and Zheng (2015) do not explain the implications of this practice. It is likely that the restriction of the availability of information is the principal reason why HFT is slower in Chinese markets, as the information set evolves more slowly. On balance, this may not eliminate the advantage in very short-term strategies practised by HF traders; however, it does slow their activity, as a scalping strategy requires at least six ticks (in this case, three seconds). On the other hand, market information will evolve more slowly, which does not entirely preclude the concept that events such as the flash crash can occur in these markets.

4.8 Introduction to national regulatory environments

This section of the chapter evaluates the nature of the regulation that is presently in place in the UK, Germany, and the US. These three countries have been selected to illustrate their different regulatory approaches. The UK takes a corporate governance approach and gives weight to standards rather than rules. In the US, much activity is in the hands of self-regulating organisations that operate at some distance from regulators. On the other hand, in Germany, regulators are a little more proactive. It is shown in this section that HFT activity is controlled by principles regarding algorithmic trading. Only in Germany has a legal Act been implemented to control HFT activity, although its provisions apply to all trading algorithms. This section also summarises the position of the EU in setting regulations within its member states. Broadly speaking, there are three structures of regulation seen in Western markets according to Lee (2011). These are government-led models (such as France, Germany, and Japan), cooperation-led models (such as the US and Canada), and flexibility models that combine elements of the two (e.g. the UK, Australia, and Hong Kong) (Lee, 2011). These approaches are explored in the following examples in the sections below.

4.8.1 Regulation in the UK

Historically, the UK has adopted a rather flexible approach to regulation; however, over time, since the 1990s, this has slowly evolved into a more government-led approach to regulation, with increasing scope added to the remit of the conduct regulator (the Financial Services Authority [FSA], which later was replaced by the FCA). This change in part was brought about due to the process of EU harmonisation. The scope of the former FSA’s powers, which have subsequently been dispersed, is shown in the table below, which is derived from Lee (2011).

Table 4.1 An overview of the UK's financial regulatory framework (adapted from Lee [2011])

	Rulemaking	**Monitoring**	**Enforcement**
Prospectus disclosure	Treasury	FSA	FSA/courts
Securities distribution	Treasury	FSA	FSA
Listing – ongoing disclosure	Treasury	FSA	FSA
Issuer corporate governance	FSA/panel	FSA	FSA
Market abuse	Treasury/FSA	FSA	FSA
Trading rules	LSE	LSE	LSE
Marketplace oversight	Treasury/FSA	FSA	FSA
Brokers – investment firms	Treasury/FSA	FSA	FSA
Clearing and settlement	Treasury/FSA	FSA	FSA

Although the source of this information is from 2011, the description seems to reflect the framework before the 2008 financial crisis. As a result of the lessons learned from the 2008 financial crisis, the FSA was replaced by the FCA and the Prudential Regulatory Authority (PRA), which have more rigidly defined remits. As regulatory activity is spread across a greater number of institutions, the Treasury serves to guide allocation of responsibility.

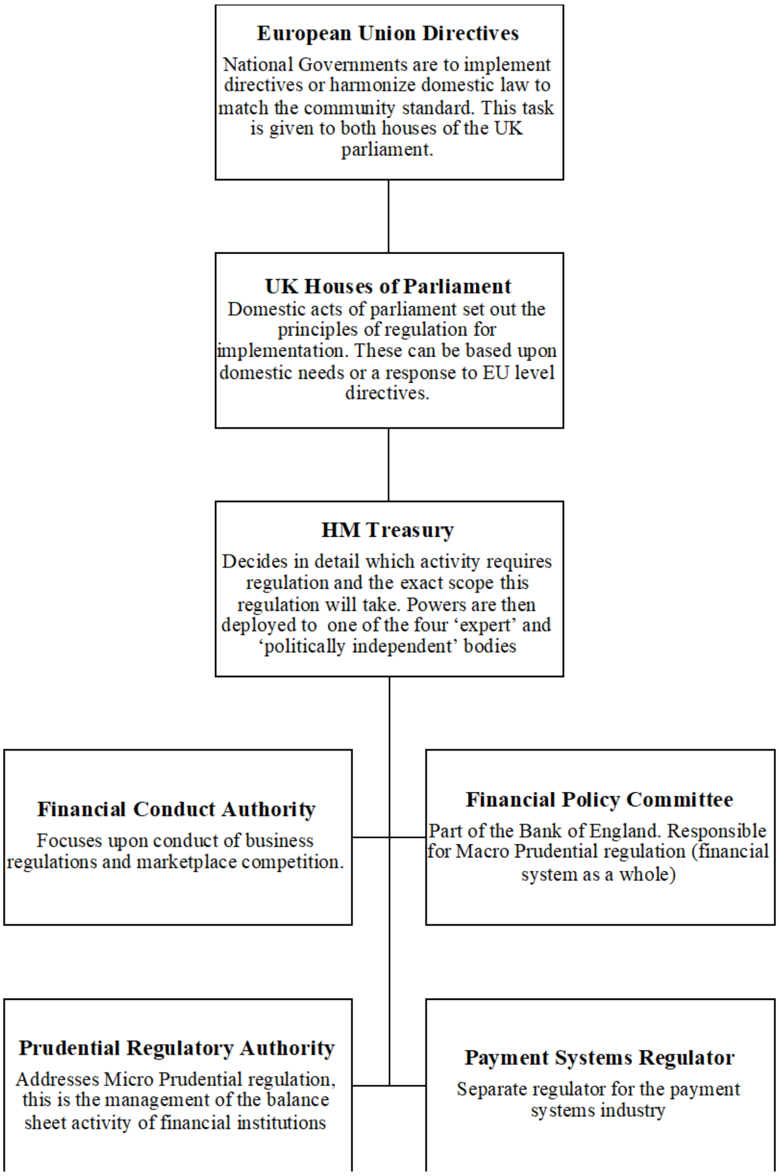


Figure 4.1 Diagrammatic summary of the UK regulatory framework (HM Treasury, 2019).

Figure 4.1 summarises the four UK regulatory bodies that were created to possess 'expert' knowledge of the marketplace as well as 'independence' from the UK Treasury and political interference. The FSA has largely been replaced by the FCA; however, its remit concerning macro- and micro-prudential regulation has moved to two new institutions: the PRA and the Financial Policy Committee, respectively. Both these bodies are under the auspices of the Bank of England and derive their independence from the Treasury in the same way as the central bank as a whole. More importantly for a study of HFT is the role of the FCA, the regulator tasked with issues of market microstructure and the efficiency and fairness of markets. The FCA's website describes the body as being the 'conduct regulator' for approximately 60,000 firms in the UK and as having standard-setting authority for around 19,000 of these firms (FCA, 2016). The remit is described as having three areas: consumer protection, promotion of competition, and upholding the integrity of financial markets. This responsibility is derived from the Financial Services and Markets Act (2000).

EU directives that have been adopted into UK law in essence remain in effect; however, no further EU directives will be adopted (Lang et al., 2017). In the post-Brexit environment, EU regulation is simply being replaced by the concept of the British common law.

4.8.1.1 Guidelines implemented in the UK

PRA Policy Statement PS12/18 concerns the use and control of algorithms. In this case, the PRA sets requirements for the corporate governance of algorithms but does not adopt a disclosure and labelling requirement as seen in the German case discussed previously. The requirements set out that a firm's governing body should allocate responsibility for the internal validation of new algorithms to a function of the firm's senior management. This approval process is required to be 'robust' and consists of a 'minimum set of risk controls' (Algorithmic Trading, 2018).

Part of the risk management expectation is to consider the effects of an algorithm's potential interaction with those in use in the marketplace and algorithms of a similar nature. The PRA policy statement highlights feedback from stakeholders, which illustrates that the level of knowledge with respect to competing firms' algorithms is rather weak. Although the PRA retorts that firms only need to consider potential interaction, it does appear that, on balance, the proprietary nature of algorithms may suggest that this is not entirely possible and that this element of risk management, in practical terms, is harder to achieve that it might initially appear. Furthermore, emphasis is given to risk associated with an algorithm's interaction with the wider 'trading system architecture'. Firms are required, as part of their supervision, to produce an internal risk management mitigation plan, this is the responsibility of the firm's senior leadership; however, they are free to call upon expertise from other operational areas within the firm (Algorithmic Trading, 2018).

The PRA issued Supervisory Statement SS5/18, which builds on the aforementioned Policy Statement PS12/18. It begins by outlining the governance requirements as discussed above, so these need not be repeated.

In addition, the PRA mandates that the governance of algorithms used in trading must incorporate (Algorithmic Trading, 2018):

- oversight of how an algorithm is used (meeting the stated purpose) and how the firm will monitor this use over time;
- a review of the continuing compliance with the conditions of the original approval;
    - This provision might sound rather vague; however, it reflects the scope for every firm to arrange this process in a different way, as the guidelines do not stipulate exact methods.
- assignment of reasonability for algorithms and the assurance that records are maintained and are up to date;
    - The PRA, at a minimum, expects each algorithm to have an 'assigned owner' who takes responsibility for the development, implementation, and use of a trading algorithm. It is the responsibility of the assigned owner to ensure the algorithm is used only as intended and designed.
- assignment of responsibility for the design and implementation of 'kill-switch controls';[27] and
    - This provision does not appear to be mandatory; however, it does appear to be a common element in algorithm design and implementation, as a kill-switch can take the algorithm out of market conditions for which it is not designed.
- implementation of a procedure to review incidents that have (or may have) involved an algorithm or where an algorithm has not operated as intended.
    - Once again, no details are given, and firms are free to implement such procedures as they see fit.

Thus, the PRA expects that the management body of a firm has both strong oversight and understanding of the operations of the trading algorithms in use in the firm. Given that the report also states that management, when reviewing an algorithm, may need to draw on expertise from other areas of its firm for guidance, management likely often lack an overall understanding of the matter. Ordinarily, those who work in the financial

[27] These controls can operate in a similar way to a circuit breaker but can also be instigated by a supervising human.

sector and hold an 'approved person status' will hold an industry qualification, such as a CFA qualification. These qualifications do not cover algorithmic trading or financial market microstructure, and this gap perhaps highlights that the responsible persons within a firm are not likely to be qualified to make judgments independent from the operational group that runs the algorithm and that contains the specialist knowledge.

A list of the minimum requirements of the management body is given in PRA Supervisory Statement SS5/18 (Algorithmic Trading, 2018):

- Traders should be instructed to understand the behaviour of an algorithm before its use, as well as its effects upon markets and liquidity.
- Traders' access to algorithms should be appropriate and subject to risk management.
    - The intent of this risk management is to account for the firm's overall risk tolerances based on client risk tolerances. The risk management must include cognisance of risk tolerance and the possibility of not fulfilling an obligation.
    - The risk management process must also anticipate the potential order flow generated, and this must match the post-order flow capacity.
- Traders' use of algorithms must be subject to the oversight of management.

In addition, individual firms are expected to create their own algorithmic trading policies. A list of minimum requirements is provided and is summarised as follows:

- The use of and purpose for using trading algorithms are identified.
- A working definition of the term 'algorithm' is given.
    - Each firm may set its own definition of an 'algorithm' – no definition is set by the PRA.
- Processes to validate new algorithms and remove existing algorithms from service are given.
- A validation and testing process for new algorithms and the periodic[28] review of existing algorithms is given.

[28] The frequency of these reviews is at an individual firms discretion

    - Thus, scrutiny of an algorithm is internal, and it is largely within a firm's individual discretion concerning to what extent an algorithm is scrutinised.
- A risk management policy that defines a series of minimum risk controls and a process of 'risk control ownership' are provided.

In summary, the PRA sets out a framework of minimum standards that apply to the use of trading algorithms and do not specifically address HFT activity. The requirements summarised above from Supervisory Statement SS5/18 relate to the corporate governance of a firm; however, very little information is given concerning what the minimum risk controls should be or what considerations should to be made.

Although some of the HFT issues are perhaps closer to the remit of the conduct regulator, clearer guidance on minimum testing standards and perhaps the use of test data in a sandbox environment could be considered here to ensure that firms' internal controls and due diligence are adequate for running algorithms that are, by design, sensitive to market conditions. An additional comment here is that, although firms are required to consider the use of kill-switch controls, these do not always operate effectively, as seen in the May 2010 market volatility event.

4.8.1.2 The role of the Financial Conduct Authority

The FCA is the designated financial conduct regulator in the UK. The FCA appears to take a more active role in the regulation of algorithmic trading and can enforce legislation. The FCA takes a firm-specific approach and a sector-wide approach to the regulation of algorithmic trading.

The FCA does support firms in providing market abuse training sessions for staff involved in institutional compliance as well as those involved in the validation of an algorithm. It is noteworthy that the FCA is providing practical steps to assist firms in meeting the governance guidance set out by the PRA, which either suggests some overlap in remit or an unclear demarcation between the roles of the PRA and the FCA.

The FCA does propose a definition of algorithmic trading,[29] derived from article 4(1)39 of MiFID II. The exact definition, as written, is given in an FCA compliance report (2018):

> Algorithmic trading:
>
> Trading in financial instruments which meets the following conditions:
>
> (a) where a computer algorithm automatically determines individual parameters of orders such as whether to initiate the order, the timing, price or quantity of the order, and how to manage the order after its submission.
> (b) There is limited or no human intervention.
> This does not include any system that is only used for the purpose of routing orders to one or more trading venues or the processing of orders involving no determination of any trading parameters or for the confirmation of orders or the post-trade processing of executed transactions.

Based on this definition, the FCA can request details from any individual firm that uses trading algorithms meeting these requirements. When the FCA makes such a request from a firm, the firm is obliged to respond within 14 days with a description of the algorithm, the strategies it deploys, and the order types involved. The FCA also appears to defer to the PRA guidance that an individual firm should set its own working definition of a trading algorithm as part of an ongoing process of identification and review of the suitability of trading algorithms in use.

The FCA also elaborates on the testing and approval process which is summarised above as described by the PRA. New algorithms or amendments to existing algorithms must pass through an internal validation process to ensure the algorithm operates as designed, exhibits the correct behaviours and correctly incorporates the firms risk preferences. In addition the FCA states that an algorithm validation process must ensure:

[29] Note the PRA does not and leaves this matter to the individual firms.

- The algorithm only behaves as intended and complies with the firm's strategy and obligations in trading.
- The algorithm is able to operate within the rules of the trading venue in which it is designed to operate.
- The algorithm must not 'contribute to disorderly trading' and must be able to effectively operate in 'stressed market conditions'. In addition, an algorithm should have an effective-kill switch functionality.

A good practice in this regard is a project-managed process for validating algorithms that are to be split into separate phases of testing. Where possible, subjective decisions should be avoided or taken on balance against other evidence or findings in order to ensure the risks are effectively assessed and that adequate risk control thresholds and kill-switch conditions are set. Records of the testing process are to be made, and they should illustrate what due diligence has taken place and illustrate how decisions 'on balance' have been made. The FCA also indicates that firms may wish to make provisions for an independent internal team (within the same firm) to review the code. This final point addresses the principal agent issue that arises from the PRA's statement, which does not make such a suggestion. It is noteworthy that this is a suggestion and is not mandatory.

Unlike the PRA, the FCA sets out some more rigid requirements with respect to algorithmic trading that apply to the algorithm and its use rather than just the corporate governance aspect. With respect to the application of risk controls, firms are expected to consider what is appropriate to protect both the interests of the firm (and its clients) as well as the wider interests of the marketplace (FCA, 2018). The origin of the more rigid requirements is the MiFID II requirements (article 15 RTS 6), which the FCA is expected to enforce in the UK given its role as the designated conduct regulator. These requirements demand the following of firms:

- Firms are to maintain market and credit risk limits.
- Firms are to set, in advance, values for the maximum order values and maximum order volumes.
- Firms should set, in advance, a message limit – the maximum rate of the submission of orders or cancellations into an exchange's matching engine.

    - This is an important consideration in preventing market abuse (for instance, the spoofing strategy), as it may limit the ability to submit and quickly cancel orders.
- Firms should consider the necessity of 'repeated automated execution throttles' that periodically require human input to keep a particular strategy running.
    - These are intended to prevent algorithms from continuing to run when market conditions are changing and becoming less suitable for a particular algorithm or trading approach.
- Firms should consider the viability of 'price collars',[30] effectively self-imposed circuit breakers that can be applied to particular assets or indexes.

Goldstein (2015) studies price collars when they were required on NYSE and found they did have a mild yet statistically significant effect in reducing intra-day volatility and may have had some effect on decoupling the equity and futures markets. When required by NYSE Rule 80A, trading firms were obliged to set price collars; presently, Rule 80A is no longer in effect (having operated between 1987 and 2007).

For the sake of clarity, firms must set risk limits, maximum order values, and message limits over units of time. Firms are obliged to consider and implement (where appropriate) execution throttles and price collars. It is likely that most algorithms are specified with an execution throttle; however, no guidance is given concerning how often human confirmation should be given, the criteria needed to confirm, the information available to the supervising human, or their suitability for the role.

Again, the guidance is unclear concerning the sensitivity of the price collars that should be set. On the one hand, this sensitivity has to be determined by the risk appetite of the firm and its clients; however, on the other hand, if the threshold is not sufficiently sensitive, there is a risk of a market volatility event or the exchange circuit breaker opening first. This situation illustrates the balance between stability and a firm's desire not to have trading halts due to a 'hare-triggered' price collar.

The FCA's compliance note illustrates the variety in terms of how firms have interpreted the Article 15 RTS 6 requirements administered by the FCA. For instance,

[30] This term does not refer to an options strategy with the same name.

some firms operate their repeated throttles by monitoring repeated orders, while others set checks on repeating orders and rejected orders per asset throughout the trading day.

In addition, the FCA also enforces regulation with respect to market abuse. These regulations apply to all investment firms and all investment activities.

The following obligations are made upon all firms:

- Firms are to establish an internal system and series of processes to monitor for evidence of market abuse and attempted market abuse.
- The monitoring of market abuse must have the capacity to evaluate each individual order, modification, and cancellation in which the firm is involved.
  - Given that this provision implies that a large amount of data needs to be evaluated, it suggests that an algorithmic approach to monitoring and compliance is expected. Where this is the case, the system should be capable of producing alerts on activity that requires further investigation and human interpretation. Records of such alerts and investigations are required to be kept for a minimum of five years and, whenever misconduct occurs, details must be shared with the conduct regulator (the FCA) 'without delay'.

Where algorithmic trading is taking place, further requirements are made:

- The surveillance system mentioned above must be automated in nature and must generate alerts to human supervisors in real time.
  - The supervisory statement uses the term 'real time'. Considering that this term may apply to HFT, it perhaps should read 'close to real time'.
  - These alerts should be designed to minimise spurious alerts.
  - Alerts should occur when indications of suspicious activity relate to other activities elsewhere within the firm.
- Records of algorithmic surveillance must be kept in a format that can be analysed internally, by the compliance department, or externally ex post.

Finally, firms must analyse the performance of algorithms as part of their post-trade monitoring obligations. Using retrospective analysis, firms must be sure that their algorithms have behaved as intended and that they have remained compliant with both the firm's risk expectations and exchange rules. Firms must also ensure that their algorithms have not contributed to disorderly trading and that they have not exacerbated conditions during a market volatility event. The FCA notes that some firms take a rather basic approach to this obligation by settings basic alerts for strategies such as spoofing, layering, and insider dealing; however, some firms do not try to detect momentum ignition strategies or quote stuffing.

4.8.2 Regulation in the EU and Germany

Within the EU member states, regulation at a national level is created within the scope of a series of 11 Directives designed to provide an element of commonality. These Directives apply in all EU member states and were the core of the UK regulatory structure for many years. However, it is now necessary to treat the UK separately, as the UK government and parliament are no longer bound in the long term by these Directives and updates thereto.

There are 11 directives that apply to capital and securities markets, which, to varying degrees, may affect the regulation of HFT activity (Hemetsberger, 2006):

1. Stock Exchange Law Directive
2. Listing Admission Directive
3. Prospectus Directive
4. Major Holdings Directive
5. UCITS[31] Directive
6. UCITS Directive (Amendments)
7. Investor Compensation Directive
8. Market Abuse Directive
9. Investment Services Directive
10. Markets in Financial Instruments Directive
11. Transparency Directive

As this chapter focuses on regulation that applies to algorithmic trading and, by virtue, HFT, it is only necessary to focus on Directives 8, 9, and 10 from the list above.

[31] Undertakings for Collective Investment in Transferable Securities.

4.8.2.1 *Market Abuse Directive: Directive 2003/6/EC of the European Parliament and of the Council on 28 January 2003 on Insider Dealing and Market Manipulation (Market Abuse)*

This Directive was made with two aims. The first goal was to harmonise the approach to regulation across the EU by providing a common set of definitions for key terms used to define insider dealing and market abuse as well as establishing a sole reasonable body in each individual member state. Secondly, the Directive exists to clarify the role of the national regulators and aid in the prosecution of market abuse where offences are committed in multiple EU jurisdictions (Hemetsberger, 2006).

The provisions of this Directive address the following issues:

- the ban on insider dealing on assets and derivative instruments.
  - Definitions of insider trading and lists of interested parties are set.
- market abuse through misinformation.
  - Guidance is given concerning the fair presentation of information and the disclosure of conflict-of-interest requirements.

4.8.2.2 *Investment Services Directive: Council Directive 93/22/EEC of 10 May 1993 in Investment Services in the Securities Field*

This Directive was designed to allow a wide range of financial service providers to operate across national borders within the EU's economic area. This harmonisation allows for a single authorisation and regulation in one of the EU member states (Hemetsberger, 2006).

The Directive provides rules on the following issues:

- guidance for the conduct of business, with respect to:
  - firms' internal controls.
  - safeguarding investments,
  - reporting and record-keeping requirements, and
  - minimum standards for investment advice;
- minimum capital regulations for deposit-taking institutions; and
- reporting and disclosure requirements between firms and national regulators.
  - MiFID rules may have superseded these requirements.

4.8.2.3 *Markets in Financial Instruments Directive: Directive 2004/39/EC of the European Parliament and of the Council of 21 April 2004 on Markets in Financial Instruments, amending Council Directives 85/611/EEC and 93/6/EEC and Directive 2000/12/EC of the European Parliament and of the Council and replacing Directive 93/22/EEC*

This regulation, while it is technically still in force, has been replaced by the MiFID regulations in 2018. Ergo, it need not be discussed. The Markets in Financial Instruments Directive (MiFID II) sets provisions regarding the use of algorithmic trading, management obligations, and the duties to a client in terms of sessing the client's suitability and obtaining best execution.

When considering algorithmic traders, MiFID requires that firms provide to their conduct regulator upon request details of the algorithms that they have in use. This information should include the strategy being used, the parameters in which the algorithm is used, evidence of suitable design elements, and evidence of testing and internal audit (Sheridan, 2017).

There are also further relevant Directives.

4.8.2.4 *Unfair Commercial Practices Directive of the European Parliament and of the Council Concerning Unfair Business-to-Consumer Practices in the Internal Market, amending Directives 84/50/EEC, 97/7/EC, and 98/27/EC (the Unfair Commercial Practices Directive)*

This Directive is introduced as having the intention of standardising the national rules across the EU member states to provide a more uniform level of consumer protection. However, it only addresses end consumers and does not make provisions for business-to-business transactions. As a result, its applicability to all undertakings in financial markets is limited.

The Directive makes provisions regarding the promotion of services:

- advertising or promoting financial products and services by financial services providers.
    - Practices can be deemed unfair when consumers, who are otherwise reasonably informed, are misled or when there has been a lack of diligence upon the part of a trader.

- Such practices can be defined further as either being misleading or aggressive. A misleading act occurs when information had not been fully and properly disclosed. The method of defining an aggressive act is not clear (Hemetsberger, 2006).
- This regulation is beyond the scope of domestic contract law and does not give basis to a claim by a consumer under the principles of tort law (the law of civil wrongs). However, EU member states are required to implement a method of their own in order to enforce penalties when unfair commercial practices are detected.

By way of clarification, this provision makes no clear indication to EU member states that all financial transactions are protected by the Unfair Commercial Practices Directive. As a result, a reasonable amount of transactions, such as those in the role of market making or institutional traders when scalping, may not be covered by the Directive. As much as HFT practice is not covered by this conduct regulation, it must be the case that national regulation is the highest level of coverage against HFT activities in EU member states.

Thus, the three most pertinent Directives of the European Parliament and Council have been summarised. They appear to provide uniformity in the protection of consumers and guidance concerning information and insider dealing. While these Directives do influence the broader financial environment, they do not directly affect market microstructure and fail to provide a direct set of principles to govern aspects of financial market microstructure. No attempt is made with respect to standardising practices in the many markets which are open to trading at the same time and may list the same assets or derivatives of another listed asset.

Overall, this omission may be seen as a missed opportunity on the part of the European Parliament and Commission due to the need to address the pressing issue of HFT practice and the wider issues of market stability and confidence in financial markets. While there is a direct consumer protection provision, there is little here to safeguard against the generation of wider spillover effects.

Thus, regulation concerning market stability and volatility and concerning the need to address HFT practices in the EU and UK exist at a national level only and will not necessarily be harmonised.

Once again, a note on the UK when considered in the context of EU Directives: Although the UK is no longer a member of the EU, its existing Directives, which have been adopted into UK law, remain in effect; however, no further EU directives will be adopted (Lang et al., 2017), although some may be mirrored in accordance with the intent of the UK parliament.

Since HFT practice is not covered by this conduct regulation, it must be that national regulation is the highest level of coverage against HFT activities in EU member states. It is not practical to make an examination of all 27 EU member states. A short discussion of only the national regulation in Germany is helpful, as its regulatory system is one of the more developed regulatory touches in terms of handling HFT behaviours.

Germany (as well as France and Japan) has government-led regulatory models. These countries combine rulemaking from the regulatory authorities with rulemaking from market infrastructure institutions, such as exchanges (Lee, 2011). In general, systemic concerns are addressed by national regulators; however, Germany empowers its institutions to address very specific issues as they see fit. Ergo, the exact market microstructure and exchange listing requirements are set by the exchanges; however, more general conduct regulation is set by regulators, unlike in the system of cooperation, as seen in the US and Canada. Here, the regulator takes a more active role in the elements of marketplace conduct and rule enforcement.

In this model, the market infrastructure is such that the institutions enjoy only limited rulemaking powers and are offered these in recognition of their 'high level of expertise' with respect to their own exchanges or indexes. This authority is held in conjunction with regulators who hold a wider remit, including financial stability and infrastructural adequacy. In Germany, this structure led to the creation of a single regulator with a large remit, Bundesanstalt fur Finanzdienstleistungsaufsicht (BfF). The logic of the large remit was to create a unified regulatory approach that could better reflect the interlinked nature and complexity of financial markets (Lee, 2011).

The German government has also produced an Act of Parliament that specifically addresses HFT rather than algorithmic trading. The title can be translated as the High-Frequency Trading Act, enacted in 2013. The key requirements are the labelling of orders and heightened licensing requirements in order to access German markets. The principle is the collection of information that allows regulators and investigations greater ability to monitor markets and investigate market abuse.

Part of this approach is seen in the German HFT Act. Here, the Act is proactive in nature, as it specifically addresses HFT rather than algorithmic trading in general. In addition, it shows the need for EUREX to adapt its rules and governance to meet the requirements; however, some freedom is afforded to EUREX.

The German Federal Financial Supervisory Authority (Ba Fin) presents a succinct summary of the challenge of regulating HFT. The summary is (German Federal Financial Supervisory Authority, 2020):

> Algorithmic trading, where orders are entered, modified, and cancelled by computers, carries various risks. For example, a high number of order entries, modifications, or cancellations within a very short space of time can overload trading systems. Algorithms may also react to market events and trigger additional algorithms as a result, which may in turn trigger even more algorithms (cascade effect), leading to an increase in volatility.

4.8.2.5 EUREX amended practices

In response to the passing of the HFT Act 2013, exchanges were required to amend their trading rules in order to maintain compliance.

The primary mechanism applies to the OTR. Traders are now subject to adequacy requirements and are required to tally their own ratio values for the ratio between ‘order and quote entries, modifications and deletions, and contracts traded’ (Peters, 2016). The OTR values are to be calculated per trader, per product over a monthly period and are subject to a series of limits. These vary according to asset classification, as summarised in full by Peters (2016). In addition to the OTR restrictions mandated, the Act implies the need for a self-matching prevention system to be in place, as where a self-match occurs, the ratio increases by one on each side. This provision appears to acknowledge that it is not common practice for individual traders to prevent crossed orders, which buttresses the remarks made on this point by Wang and Zheng (2015).

Prior to these OTR restrictions, the German Exchange Act 2014 implemented requirements for the definition and labelling of ‘trading algorithms’ with the intent of limiting the potential for market abuse arising from HFT activity (EUREX, 2013). Exchanges are required to incorporate into their own venues trading rules regarding the 'labelling of orders generated by algorithmic trading' (EUREX, 2013).

Under the HFT Act, a trading algorithm is defined as being a series of calculation steps that lead to a sequence of instructions 'with a finite length', which constitutes a trading strategy consisting of a series of orders. In addition, it is noted that this practice does not require a constant human presence, or at least, constant human input.[32] Based on this definition, the above identification obligation is applied.

Each individual algorithm is required to be uniquely labelled and a label applied – regardless of the purpose of an order (i.e. to trade or to assess price or depth). The label takes the form of a unique identification key, which identifies the processing algorithm and illustrates the origin of an order; the rationale thereof is to provide accountability and to allow for more straightforward investigation of market abuse incidents. A consequence is the requirement to disclose some information about algorithms without the need to disclose the proprietary element of an algorithm.

4.8.3 Regulation in the US

Much of the present regulatory framework in the US has its origins in the reforms following the 1929 Wall Street Crash. The first federal statute in the US that made country-wide provisions rather than state-level provisions was the Securities Act 1933, which addressed conduct on stock exchanges and established the provision of regulatory oversight (DeBedts, 1964). This Act preceded the Securities and Exchange Commission (SEC) in 1934 taking responsibility for financial markets – into the hands of a specialised regulator. Previously, the Federal Trade Commission held this responsibility and was limited by a political environment that embodied laissez-faire principles.

This event was followed by the Securities Act 1934, which remains in use to the present day, and recent convictions have been made against this statute. The 1934 Act is based on a principle of 'truth in securities' (DeBedts, 1964, p. 57). It provides for the federal listing of all US exchanges within a regulatory system designed to provide

[32] The source uses the phrase 'while a continuous human interference is not required for this purpose', which implies that a human presence in a monitoring role is expected.

very little regulation though direct statutes. Rather, the idea was to leave the details and rules for exchanges to the discretion of a regulatory agency that can adapt regulation to meet the needs of future market conditions and changes in market practices. In practice, this is the basis of the Directives of the SEC and their basis in law. It is also the 1934 Act that provided for industry representatives from stock exchanges to engage with the SEC to provide technical assistance (DeBedts, 1964).

This is the basis of the present regulatory structure, tempered with some deregulation and additional regulations and the increased oversight of the Dodd-Frank Wall Street Reform and Consumer Protection Act 2010.

The US's (and Canada's) regulatory approach is built on a cooperation model (Lee, 2011). This structure places emphasis on self-regulation, and the term self-regulating organisation (SRO) is incorporated into the regulations and guidance produced by the SEC. Enforcement of the regulations rests with the internal structure of an SRO, and it is expected that a firm will invest significant resources in self-compliance.

Lee (2011) summarises the overall regulatory structure in a table which adapted and presented below. Figure 4.3 illustrates the wide spread of the SRO concept and as of 2006 the presence of only one regulatory body with a wide remit. This table is presented to enable comparison with table 4.1 depicting the UK regulatory structure.

Table 4.2 Overview of the US's financial regulatory framework (2006). Adapted from Lee (2011)

	Rulemaking	**Monitoring**	**Enforcement**
Prospectus disclosure	SEC and SROs	SEC and SROs	SEC and SROs
Securities distribution	SEC and SROs	SEC and SROs	SEC and SROs
Listing – ongoing disclosure	SEC and SROs	SEC and SROs	SEC and SROs
Issuer corporate governance	SEC and SROs	SEC and SROs	SEC and SROs
Market abuse	SEC and SROs	SEC and SROs	SEC and SROs
Trading rules	SEC and SROs	SEC and SROs	SEC and SROs
Marketplace oversight	SEC and SROs	SEC and SROs	SEC and SROs
Brokers – investment firms	SEC and SROs	SEC and SROs	SEC and SROs
Clearing and settlement	SEC and SROs	SEC and SROs	SEC and SROs

The US does also have the Commodity Futures Trading Commission (CFTC); however, the CFTC has a very limited focus on options, swaps, and derivative instruments as well as digital currencies. This body is highly specialised, as it largely appears to work alongside the SEC within the SRO structure of regulation. Lee (2011) does not mention this point; however, the body has existed since 1975.

Securities issuers can cross-list assets on multiple markets in the US. Furthermore, within the EU, assets can be traded across national borders, as markets are mostly open at the same time (Morelli, 2017). It is possible that an exact dual listing may exist; however, similar issues exist between assets and derivatives of an asset traded in different marketplaces. Evidence of this form of derivatives coupling is discussed by Goldstein (2015).

In the US, a joint committee consisting of representatives from the SEC and Commodity Futures Trading Commission (CFTC), the Joint Advisory Committee, has been established to provide coordinated responses to perceived issues with respect to market micro-structure (Morelli, 2017). Where a recommendation is provided, the committee has no authority to enact it; however, the SEC or CFTC may adopt the recommendation and incorporate into their own regulations. This is pertinent because, in order to implement trading controls such as batch auctions (batch auctions control the supply of an asset over time and may restrict activity that uses latency-based strategies), wider consultation would likely be required before a regulator can approve a proposal. Ergo, it may not be the case that the SEC alone would take this initiative. In addition, regulatory coordination is needed to alter or set standards with respect to tick sizes, trading fees, and message limits (Morelli, 2017).

With respect to harmonising the regulation of algorithmic trading and HFT between countries, Morelli (2017, p. 228) argued that this would be both ‘contentious and impractical’, and it would require an international body to coordinate, which would supersede national regulations and legal processes. Furthermore, a series of globally standardised regulations would not be well placed to account for national variations in market structures. In addition, as highlighted in Chapter 2 of this thesis, there is a significant variation in venue characteristics, such as the signal-to-noise ratios. On balance, it may be helpful to standardise some of the governance concepts concerning algorithms across countries. In the EU, this is partly the case, as illustrated previously.

However, the variation in venue characteristics is likely the reason why the EU Directives go no further than this.

Woodward (2017) notes that regulators may opt to exercise 'appropriate restraint' by which an algorithm may deliver misinformation to the market; however, the effects thereof are left to the marketplace, as the strategies of 'smart traders' who attempt to trade their information into price, as if correcting the effects of 'noisy traders' as outlined in Black's noise trader model (Black, 1986). In addition, regulators must develop the ability to distinguish, in their analyses of trading records, between legitimate and vexatious practices (Woodward, 2017). If regulators cannot easily conduct a retrospective analysis, then how conduct regulators can enforce regulations can be called into question.

In the US, HFT activity is regulated by the SEC, CFTC, and Financial Industry Regulatory Authority (FINRA). SEC regulation SCI outlines the following:

- Some institutions are able to self-regulate, typically stock exchanges, options exchanges, and clearing houses.
- Data should be provided to the consolidated market data processor, previously outlined in the regulation outline as the US consolidated tape.
- This regulation can apply to 'alternate trading systems', also known as 'dark pools', where volume thresholds are exceeded.

A firm covered by regulation SCI is subject to governance requirements that seek to ensure 'fair and orderly markets'.

Legislation upheld by regulation SCI is the US Securities Exchange Act 1934 and minor additions thereto.

Some companies have also been affected by the introduction of SEC rule 13h-1, which requires large traders to meet certain standards of recordkeeping and reporting, and it provides for the monitoring of certain brokerage activity. The rule does not target HFT but rather seeks to allow the SEC to gather information on participants who substantially contribute to market volume and price impact.

Following the May 2010 flash crash event, the Securities Industry and Financial Markets Association (SIFMA) proposed the implementation of acceptable price bands (it is not clear exactly how this may be implemented) and a system of market alerts to draw the need for additional liquidity to the attention of traders (Woodward, 2017). At the time of writing, this proposal is with the SEC for consideration. Although it is likely intended to improve retrospective analysis of specific events, it is the opinion of Keller (2012) that, given this disclosure, the scope for improved analysis remains poor. On balance, a consolidated tape or consolidated audit trail is more helpful to those undertaking retrospective analysis.

4.8.3.1 US enforcement activity

There have been a few cases in the US where enforcement action has been taken against traders involved in HFT. While the US does not have a common-law system, these cases are useful examples of the involvement of the SEC and what action could be expected in the future.

Athena Capital Research LLC

Athena Capital Research LLC was sanctioned in October 2014 by the SEC for using a series of aggressive HFT tactics. For a period of approximately six months, the company used an algorithm that had been codenamed 'grave' in order to execute a 'marking-the-close' strategy that was designed to create an orderbook imbalance (call and put orders that are not equal). Following an investigation by the SEC's Division of Enforcement, the SEC charged the company subject to the Securities Exchange Act 1934, and a fine of US$1 million was imposed, although Athena denied the allegations (Woodward, 2017).

Haim Bodek

Bodek had gained experience in electronic trading at Hull Trading (an early adopter), Goldman Sachs, and more recently, UBS. He then went on to run his own firm, Trading Machines, beginning in 2006. While running his own algorithmic trading company, Bodek encountered diminishing returns on his algorithms (beginning around 2009) and until 2011 did not have a sound understanding of the reasons for this when Trading Machines ceased to operate in January 2011. Anecdotally, at this time, Bodek felt 'something in the market was rigged' (Patterson, 2012).

Bodek latter alleged that some traders had developed an ability to 'jump ahead' in the queue for execution, and, as a result, the exchanges gave preferable execution in terms of speed and best price. This claim is elaborated in a book published by Bodek (2013).

However, on balance, there is no supporting evidence for it. While the idea may not withstand academic scrutiny, it is known that orders of the 'hide-not-slide' type are not cancelled and re-entered when markets lock, which would have the effect of rearranging the order in the queue for execution. Given that this situation can occur, it is possible to conclude that Bodek did encounter a permutation of this order type, in which case, it would be possible to perceive that some orders have 'jumped ahead'. On the other hand, exchanges do not make visible details of the queue for execution; ergo, it is unclear concerning how Bodek reached this conclusion, as he had no information available on other order types in the market. While he may suspect it, he cannot presently prove it.

The abovementioned allegation was the basis of a complaint made by Bodek to the SEC in 2011, and the SEC was only able to respond to the complaint in January 2015 after an investigation that lasted four years. Thus, the SEC may have had some difficulty in understanding the nature of the order types in use and the general market microstructure. The SEC found, in the course of the investigation, that the trading venue, controlled by Direct Edge, did not provide all traders with the necessary information about all order types approved for use on the exchange and did not make available information regarding the ability of order types to affect priority for execution. The SEC found that this information had been disclosed to some traders (mostly large HF traders); however, it had not been shared with all members. The SEC concluded that a violation of sections 19(B) and 19(G) of the Securities Exchange Act 1934 had taken place and fined the EDGA and EDGX exchanges (who had subsequently acquired control of Direct Edge) US$14 Million (United States of America vs EDGA Exchange, Inc., 2015).

This incident illustrates the potential difficulty in assessing events using a retrospective approach, as it required four years for the parties involved to achieve an understanding of the events that had taken place. The core element of misconduct on the part of the exchange in this matter raises questions about the suitability of the SRO role within the cooperation model of regulation, as the SEC, in its role as conduct regulator, was not involved in the governance of the exchange until they became obliged to respond to a formal complaint made by Bodek. This example perhaps makes the case for conduct regulators to make periodic reviews of firms' internal regulatory processes to ensure that the compliance role is sufficient and that firms monitor activity appropriately. In this case, the rules-based capitalism element of the marketplace had been removed by the entity controlling the marketplace. The crux of this issue appears to be that the firm was not able to meet its obligations with respect to its self-regulation and may have benefitted from a more active form of supervision from the conduct regulator. In short, the exchange did not comply with its own rules from July 2010 onwards, and this lack of compliance was not detected.

Although the US authorities practice what may be considered a light-touch regulatory approach, it is clear from the above cases that they also lack core knowledge of activity taking place on markets for which they are responsible. On the other hand, the US has been quick to open investigations and take the need to uphold the integrity of markets seriously. This is especially visible in the case of the SEC ruling against an exchange operator. This slow-to-learn but quick-to-prosecute approach is a poor reflection of the idea that a firm can self-regulate in the first instance. The US regulators have the opportunity to require firms to disclose more information to them, which would make future investigations much quicker and draw out the learning points much faster.

4.9 Compliance

A common theme in the conduct regulation described above in all the three cases above is the role of compliance and the duties of the compliance function. Financial firms need to give weight to compliance as part of their adherence to regulation but also to the minimisation of 'compliance risk', the financial loss or reputational damage caused by a failure to meet regulatory standards. Due to the risk of fines, compensation claims, and reputational damage,[33] firms have an internalised incentive to minimise compliance risk by proactively implementing an effective compliance function. In addition, compliance can be described as an organisational culture that needs to be part of the working practices of all staff in addition to the specialist staff employed within the compliance function (Schilder et al., 2005).

The Basel Committee on Banking Supervision (a division of the Bank of International Settlements) provides ten principles of compliance, which are helpful for

[33] This risk applies especially when a firm is involved in consumer finance, where goodwill (or, more precisely, brand image) is a more pressing concern.

understanding the role of compliance in a financial services firm. These principles are summarised below (adapted from Schilder et al., 2005):

1. A firm's board of directors is responsible for overseeing the reduction and minimisation of compliance risk through the establishment of a permeant and well-resourced compliance function.
2. A firm's senior management is directly responsible for the management of compliance risk.
3. A firm's senior management is reasonable for communicating compliance policy within the firm and actively ensuring that policy is observed.
4. A firm's senior management has an obligation to implement a permanent and effective compliance function. This function rests alongside the board of directors' obligations.
5. The compliance function should be independent from other internal functions and should have its own formalised status within the firm.
6. The compliance function should be well resourced and able to carry out its responsibilities effectively.
7. Appropriate compliance duties may need to be carried out by other business functions, such as when specialist knowledge is not held within the compliance function. If this is the case, the responsibilities must be clearly set out by formalised arrangement.
    a. The guidance does not address the potential for a conflict of interest when a group may validate or monitor its own work. This is possibly a concern in terms of algorithmic trading and HFT.
8. The internal audit function should periodically review the scope of the activities of the compliance function. The internal audit function must be separate from the compliance function.
9. Appropriate firms must comply with regulations and laws in all countries in which they operate. If firms are operating using a subsidiary model, this does not entirely absolve the parent company from this requirement.
10. Compliance should be considered part of a firm's core risk management activity.

This list of principles is drafted with investment banks in mind rather than traders or exchanges; however, the concept of compliance is somewhat generic across the financial services industry as compliance addresses the adequacy of a firm's actions. As these principles are set out by the Basel Committee on Banking Supervision, they are for guidance only and cannot be taken as mandatory, unless a national legislature chooses to enact them as law. Note that these principles reflect the PRA requirements set out in the UK very well.

It is possible for these guidelines to apply in multiple jurisdictions given their partially generic nature; however, it must be noted that the principles are provided along with further explanatory notes, which are not replicated here. An interesting omission in the guidance relates to compliance functions being able to outsource some tasks to other business functions. Hypothetically, it is possible for those who design a trading system to be responsible for its internal validation, which may especially be the case within small firms that are not well resourced.

4.10 The role of active regulation and monitoring

In order to illustrate the concept of passive regulation and the concept of monitoring firms' compliance with regulations, a two-player 'inspection game', as given by von Wangenheim (2004), with a slight adaptation, can be used.

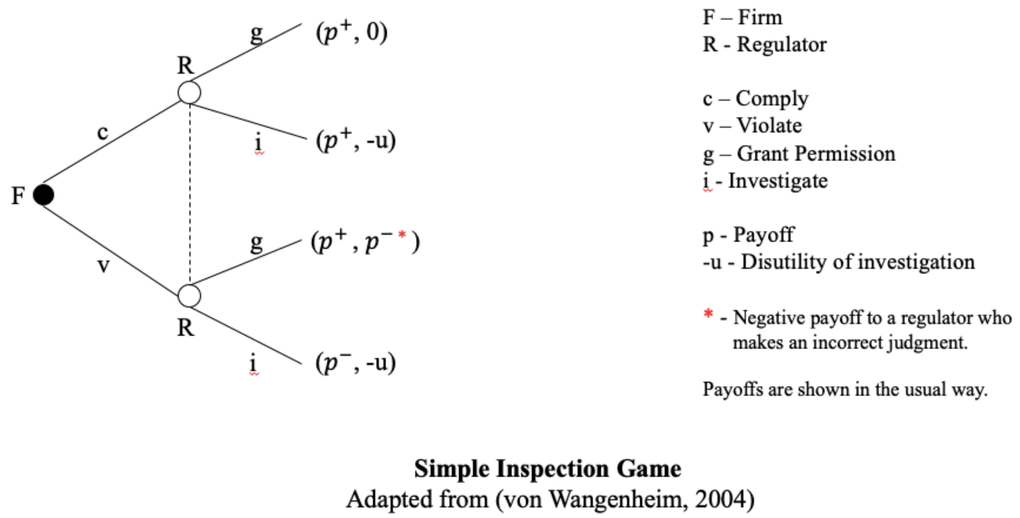


Figure 4.2 A simple inspection game

Two players – a trader (the firm) and a regulator – exist in this game. Firstly, the firm makes an undertaking and approaches the regulator. This undertaking may be in line with regulations or otherwise. Not knowing the firm's course of action, the regulator will choose to expend effort to inspect the firm's activities or to take the firm at its word.

The solution is reached by mixed strategies and best responses.

According to von Wangenheim (2004):

Equation 31

$$p_v^+ > p_c^+ > p_v^- > p_g^- > -u > 0$$

The solution by means of the best-responses method is marked in bold, so a pure strategy Nash equilibrium exists here.

Without numeric payoffs, calculating an exact mixed strategy response is not viable, and von Wangenheim (2004) does not attempt this; rather, they illustrate best responses changing as the probability of compliance and investigation change.

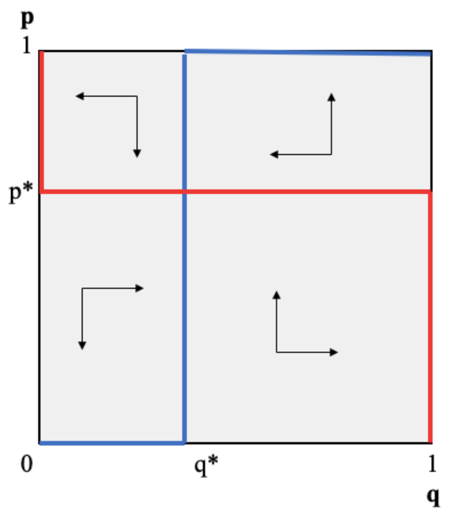


Figure 4.3 Best responses in the inspection game

Blue denotes the probability of compliance and red denotes the probability of investigation. Across the probability space, the dynamics are indicated by black arrows. Here, p denotes the probability of compliance and q denotes the probability that compliance would be verified.

A key point from this illustration (Figure 4.3) is that the probability of a firm's exact compliance with regulations decreases as the probability that the regulator will investigate its activity decreases. The mixed-strategy diagram illustrates these dynamics. It is also clear from the lack of a pure-strategy Nash equilibrium that this is a rather unstable relationship, meaning that enforcement of regulation has an effect in preventing activity that the firms are not able to prevent through internal means. This finding lends some support to the idea that instead of being too big to fail, a firm can also be too big to manage.

Excessive risk taking and market misconduct is caused by poor internal oversight due to the scale of operations, which reduces the directors' and the senior management's ability to exercise their required oversight with respect to market discipline and internal validation. This issue is compounded by a regulatory structure that emphasises strong internal governance rather than external oversight. This is indeed the structure in the UK and the US. Existing literature applies the 'too-big-to-manage' concept to banks, particularly mega-banks or 'systemically significant banks'; examples include Cerpa Vielma et al. (2019) and Kress (2019). A solution to the 'too-big-to-manage' problem was suggested by Kress (2019). The idea is to avoid disbanding larger banking groups into smaller institutions. Rather, the idea advanced is for regulators to be empowered to require firms to divest a particular operating area, capitalised by the same shareholders with new management, only where oversight can be shown to be lacking; in other words, action only need be taken where the 'too-big-to-manage' problem is felt to exist. This solution does not directly seek to disband large firms and reduce single-market concentration; as a result, it is perhaps an ordoliberal form of intervention. The suggestion does, however, require that regulators take a somewhat active approach. With regard to algorithmic trading, regulators would require a means of auditing firms' internal validation processes and investigating the role of an individual algorithm following a market volatility event. The German HFT Act provides a framework for the latter.

Taken alongside the evidence of the simple inspection game, there is a clear incentive for firms to violate rules and avoid meeting standards when inspection is an exception rather than an expectation. Combined with the scope for poor oversight, by way of the 'too-big-to-manage' problem, a possible lack of direct supervision, or otherwise, it is possible that algorithms may be used in strategies that are not in the interests of wider market discipline and stability.

In his treatise on gaining and keeping power, Niccolò Machiavelli poses the question, 'Is it better to be feared or loved?'. For a leader to be trusted, they need to 'go about things coolly, cautiously, and humanely', but if the leader is too trusting, they make themselves 'unreliable' (Machiavelli and Bull, 2003). Ergo, a leader needs to rely upon a mixture of fear and affection; in the case of the conduct regulators, they probably enjoy neither. This is a reminder of the inspection game found in von Wangenheim (2004) and the importance of monitoring.

4.11 Notes on the applicability of caveat emptor

DeBedts (1964) raises the possibility that statues and regulations can leave ambiguity or in cases fail to provide a direct provision as to what is and is not legal. In such instances the concept of common law is needed to establish what the law is. In this UK this is a long established procedure however this can happen in more limited circumstances in Mainland Europe and the USA. This section explores the direct regulation and scope for common laws applicability in the UK and the US.

This is highly interesting as if a common law principle known as 'caveat emptor' can be said to hold, then it would be the case that prosecuting market misconduct which has had an adverse effect upon others would be near impossible.

The aftermath of the 1929 financial crisis led to a significant change in regulatory approach in the US, which took the form of the Exchange Acts of 1933 and 1934. Beforehand, Huston Thompson, the chairman of the Federal Trade Commission, postulated that the change in approach would cause one of two things. Either it would '[change] the ancient doctrine of caveat emptor to "Let the seller beware" or "Let the seller also beware"' (DeBedts, 1964, p. 33). These quotes taken from a letter written by Thompson to President Franklin Roosevelt, dated March 1933; the letter does not seem to be freely available. Roosevelt is believed to have sided with the principle of 'let the seller also beware'. In either case, there is an implication in the belief that the term caveat emptor would apply. Ergo, at the time of the Exchanges Act 1934, the foundation of the present US regulatory system, there was a belief in the principle of caveat emptor.

Caveat emptor is an expression used in common law, typically translated from Latin to mean, 'Let the buyer beware'. This term is known to have been in use in the US and in the UK in the common and contract law. It is not immediately clear if this concept could ever be used as a defence in cases of market malpractice; as a result, it is necessary to perform a short legal review to clarify this point and also achieve a more exact definition of the term. The need for this review and its originality suggests the term is not in common use.

The origin of the term caveat emptor appears to be found in UK common law, outlined in an appeal case heard in 1604. The case of *Chandelor v Lopus* [1604] centred on the sale of a bezoar stone (which at the time was believed to have healing properties). Chandelor (a goldsmith) sold, for the sum of £100, what he claimed to be a bezoar stone[34] with an implied warranty to that effect. Later, the buyer, Lopus, discovered the stone was not in fact a bezoar stone and claimed that he had been deceived, thus suing for the return of his £100.

When the case was first heard, it was believed that Chandelor, who, as a goldsmith, would likely have known the stone was not a bezoar, did misrepresent the stone. Having pleaded not guilty, the court ruled in favour of Lopus (the buyer) (Moglen, 2020).

The case was subsequently appealed to the Exchequer Chamber. All but one of the judges held that, as an express warranty was not established, the burden of the risk was with the buyer (Lopus) rather than the seller. One judge (Anderson) in the case dissented, arguing that there was an element of dishonesty in the case; nevertheless, Anderson supported the overall verdict. The case was not further appealed to the House of Lords (Moglen, 2020).

The basis of the verdict is a belief that 'everyone, in selling his wares, will affirm that his wares are good' or that the terms of the sale were fair. This appears to mean that the judges believed that it was normal practice for a trader to favourably present a product or service and a case of deceit is only possible when a clear case of false warranty is made (rather than a false affirmation). Ergo, in most cases, the onus is upon the buyer to sess the quality of goods and the fairness of the price of these goods.

This set a precedent in UK law, which appears to have held until the 20th century as a concept in tort and early contract law. Cases that reverse the principle in *Chandelor v*

[34] This is a solid mass that forms on the digestive tract of animals and humans. At the time of this case, they were somewhat believed to have healing or magical properties. This point does not have a bearing upon the findings of the case.

Lopus [1604] have variations that illustrate the common law as it stands; ergo, it is necessary to summarise three cases.

Firstly, *Oscar Chess Ltd v Williams* [1957] concerned the sale of a Morris 10 motorcar. Oscar Chess Ltd was a firm of motor dealers and represented the car for sale as a 1948 model based on the logbook supplied to them with the vehicle. Williams, upon buying the car, subsequently discovered that the car was in fact a 1939 model and sued for breach of warranty. The court found that the defendants had not entered into an obligation with respect to the age of the car; however, the misrepresentation of the age of the car was an honest mistake as the car had been bought into the trade with a false logbook (Wilson, 2012). This judgment does not exactly imply caveat emptor; however, two key elements of the decision are the lack of deliberate information on the part of Oscar Chess Ltd and the lack of an explicit warranty given to Williams.

This case is contrasted with *Dick Bentley Productions v Harold Smith Motors Ltd* [1965]. In this case, Harold Smith Motors Ltd sold a car claiming it had driven 20,000 miles since a new engine and transmission had been fitted; in fact, the car had driven 100,000 miles, and the car proved unsuitable. At trial, the court ruled that the motor dealer held specialist knowledge and should have known the car had driven a greater number of miles. In addition, as the fundamental description of the good was a contract, the court ruled in the buyer's favour (Wilson, 2012). Here, the concepts held in *Oscar Chess Ltd v Williams* [1957] were reversed, as an applied warranty was acceptable and specialist knowledge was accounted for.

The *Oscar Chess Ltd v Williams* [1957] case illustrates that a court considers honest mistakes. It is also similar to the rule of caveat emptor in that a fundamental of sale was not seen to have been warrantied. The situation was the opposite in the case of *Dick Bentley Productions v Harold Smith Motors Ltd* [1965], which was tried eight years later. Here, the concept of fundamental of sale is not required to be explicitly warrantied and the concept of specialist knowledge is introduced, which was not appropriate in *Oscar Chess Ltd v Williams* [1957].

As an additional note of clarification, in UK law, it is not possible for a party to sue for damages in cases in which misrepresentation can be shown to have been innocent (Wilson, 2012).

The FCA expresses a similar view: that the asymmetry in knowledge means that those who sell financial products have a responsibility to assess the suitability of the consumer rather than rely upon the ability of the purchaser (O'Loughlin, 2013). This point is especially important, as behavioural economics has illustrated that rational choice is seldom observed.

4.11.1 Caveat emptor in the US

The term caveat emptor appears to have existed in US contract law and is incorporated in the case of *Seixas v Wood* [1804], which derived its decision from *Chandelor v Lopus* [1604]. It appears to have been common for US courts that lacked precedents to refer to UK common-law cases, although this practice became less common over time as US-based precedents have become more numerous and the scope of statute law (which takes precedence over common law in most US states) had become broader (Pope, 1910).

However, this concept in the US was later overturned by the case of *Hawkins v Pemberton* [1872]. In this case, Pemberton sold bottles of blue vitriol to Hawkins that transpired to contain an inferior product. In this case, Pemberton argued that he had made no express warranty (a defence reliant upon caveat emptor); however, this argument failed, as the court ruled that an affirmation about the description of a good can be seen as a warranty. This rule became the basis of the Uniform Sales Act 1906 that was adopted in many US states. However, the Act focuses on goods with physical characteristics; ergo, it is not clear concerning to what extent it applies to fictitious commodities such as financial products. The Act does make provision for the adequacy of price; however, its full protection for goods sold as described applies to minors and those who are not able to make decisions alone.

The concept of caveat emptor was thus not widely in use around 1936, when De Bedts was published. To the best of the author's knowledge, no US statute implies caveat emptor is still a valid legal concept. Quinn Mills (2002) illustrates that, even in well-regulated markets, the intrinsic profit motive of the financial sector can lead to activity that may be legal but not in a client's best interests. Ergo, while the principle of caveat emptor may have no basis in law, it is a reasonable principle in finance, where information asymmetry exists. This observation illustrates that a firm that is proven to have manipulated market prices through the use of illegal strategies does render itself legally liable for loss imposed on other parties, which also implies that, in law, traders have the right to expect that their orders are matched with other orders that are entered at a rational price rather than a price intended to induce momentum.

4.12 Summary

Previous chapters of this thesis have outlined some of the HFT strategies that are considered harmful to market quality and, from the perspective of firm conduct, may need to be restricted. Examples of these strategies are spoofing, sounding, pinging, momentum ignition/wash sales, and front running. Other aspects of HFT within the market microstructure that are more questionably vexatious include rapid liquidity withdrawal and associated toxic arbitrage. Furthermore, the use of algorithms in very short-term scalping strategies may have a detrimental effect upon market stability, although this point is less clear.

In the case of spoofing, although legal sanctions are in place against it, this practice can be further limited through the use of maximum OTRs, which can be built into a firm's algorithms if regulations require this. This practice would be a mechanism to disincentivise this form of market abuse rather than rely on the threat of legal sanction alone. As explained above, when a firm believes that investigation is unlikely, it has less incentive to comply with regulations;[35] ergo, the legal threat alone may not be effective. In addition, the use of a maximum OTR could limit a firm's ability to use pinging tactics to sess orderbook depth. This does not appear to be an illegal tactic in some jurisdictions and, on balance, is perhaps not the most pressing concern to most regulators.

On the other hand, momentum ignition strategies (known in the trade as wash sales) are a pressing concern, as such action seeks to push prices away from the fundamentally correct value. A firm that does this breaks national law in many localities, and, in countries that follow UK common law, such a firm would have no protection under the old caveat emptor doctrine. In this instance, algorithms are subject to internal governance requirements, and use thereof should be controlled by the compliance unit of the firm. Chapter 3 of this thesis provides a model for identifying

[35] This position calls into question the concept of 'appropriate restraint', and regulators may need to consider alternatives thereto.

first movers, which may be of use to regulators in retrospective analyses of market data to determine the origin of an ignition event. If used in a retrospective analysis, the model may help to provide a robust case against a firm that engages in momentum ignition and increase the credibility of the threat of prosecution. Once again, the incentive for firms to commit misconduct decreases as monitoring increases. Perhaps in the future, the use of RegTech may assist in scaling up analyses of price movements.

Concerning problems that are related to interaction, firstly, there is the rapid withdrawal of liquidity, which in some cases, can prevent slower traders accessing liquidity provided by faster traders (lower latency/higher frequency). There is an argument that separate venues for slower and faster traders solves this issue; however, such a separation would also create issues of a lack of liquidity in markets, and this problem would conflict traders' best execution obligations in the cases when they are trading on behalf of a client as a broker. It seems this idea has gained little traction in practice. On the other hand, imposing a minimum order resting time may slow the lower-latency traders and may also serve to keep their OTR lower. It would slow the withdrawal of orders and would go some way towards reducing toxic arbitrage situations, although a long resting time would have greater effect. The question of what is 'fair' arises again. In addition, the minimum order resting time would likely not solve all latency mismatch issues and may be harmful to the liquidity provision of scarcely traded assets.

A compromise may exist in applying separate standards for traders who are designated in their venue as a 'designated market maker' or 'market specialist'. Here, it is difficult to elaborate, as each exchange has different characteristics, so the details would be for the individual exchanges to decide. As most designated market makers, in equity markets, are known to be HFT systems, perhaps a modest order resting time against only the designated market maker may seem attractive to slower algorithmic traders. However, this would not be attractive to the market maker, as it too is an entity with a profit motive, and having its orders restricted compared to those of other traders would cause them a great disadvantage due to other HF traders' scalping its positions. Although, transaction taxes would reduce the use of scalping over short periods in general by reducing profitability, which raises the question of exchanges bringing liquidity provision 'in house'. However, this situation would be a significant change to exchanges' working practices, and, at present, no research suggests this rather drastic solution.[36] Keller (2012) did give credence to the idea that designated market makers should be identified and should be subjected to specific exchange market-making rules.

Finally, there are known issues in terms of how traders use order types to gain advantage over one another. In principle, if all traders in a venue have access to all the approved order types in the venue, traders should be able to determine how to interact with one another. Cases such as Haim Bodek's issues using Direct Edge seem to be a failure of exchange governance rather than a fault of HFT in general. This point reflects critically upon the concept of the SRO, championed in the US regulatory structure. With oversight, once again, market discipline and the governance of firms is likely to improve.

An additional key concern raised throughout this thesis is market volatility. Adherence to the algorithm design and validation requirements (for example, the FCA requirements discussed above) should minimise the effects of algorithms in volatility events. In addition, exchanges can implement circuit breakers as a backup protection; however, it is perhaps more desirable to hold as much stability from the algorithms as information allows for. As a result, governance requirements are a better solution than the use of circuit breakers alone.

The final consideration here is the cost of increased oversight. In the future, it may be possible to apply regulatory technology (RegTech) to increase the frequency of oversight and reduce the costs thereof. An algorithm-labelling requirement, similar to the one that is in place in Germany, may assist in the use of algorithmic intelligence to regulate markets. A similar proposition is made by Kavassalis et al. (2018). Little research exists on the implications of RegTech against HFT or algorithmic trading; as a result, it is not possible to exactly determine whether the current state of technology can support this level of analysis. Once again, this is an area that requires further (and perhaps interdisciplinary) research. Despite this limitation, AI provides an interesting opportunity to develop better techniques for detecting practices like momentum ignition. This is important as there is some evidence to suggest that the FCA cannot detect this activity effectively – even in retrospect.

[36] This is a good area for further research and could be a PhD thesis of its own.

4.13 Conclusion

National regulators, lawmakers, and those responsible for trading venues may all be interested in ways to regulate HFT activity. These methods will vary according to the country and regulatory structure. For instance, in the UK, much scope rests with the FCA, while in the US, more oversight and learning points needs to be taken by individual exchanges. This chapter has summarised the variances in general regulatory structures and outlines the existing regulatory tools at the disposal of those with regulatory authority. In addition, the concept of ‘fair’ regulation and ‘fairness’ in market outcomes has been evaluated as far as is reasonably possible. The primary motivation for this thorough review has been to collate what interventions are currently in use and what evidence exists with respect to their effectiveness. The second aim has been to set these interventions against ‘fairness’ in markets that contain more asymmetry than ever before.

On balance, measures do exist that form a theoretical perspective, and they can reduce the incentives to operate low(er)-latency trading systems. These measures may be effective in reducing instances of ‘toxic arbitrage’; furthermore, regulatory technology offers opportunities for improved monitoring of market conduct. This means that less emphasis needs to be placed upon firms’ self-compliance with frameworks in light of greater monitoring of traders’ actions. However, there is a need to be ‘fair’, and policymakers desire to maintain trust through more intensive monitoring practices. This work highlights the complex balance needed between the interests of multiple stakeholders. Further research is needed to define ‘fairness’ in regulatory applications to serve as a foundation for the establishment of regulation that balances the needs of multiple stakeholders.

5 Policy Recommendations and Conclusions

This section summarises some key points and sets out a series of learning points (LP) for consideration based upon the finings in each of the three previous chapters.

Chapter four has sought to evaluate the merits of various possible interventions that can be applied to limit the ability for HF traders to use illegal strategies or to limit the effects of market volatility events to which HFT may contribute. It has been shown that speed bumps in trading-type interventions are not effective, as they do not remove the incentive to be the first in the queue for execution. However, setting a minimum order resting time may change the strategy set deployed by traders, in essence, forcing them to slow down due to their need to trade profitably. After all, some orders are placed in the knowledge that they can be updated or withdrawn promptly. The implementation of minimum order resting times would also be consistent with the ordoliberal approach, as all traders could be made subject to these limitations in a particular venue, although, of course, some would be more affected than others. This highlights the key element of the ordoliberal approach outlined, as it seeks to balance the needs of different trader types rather than simply target those who others identify as the troublemaker.

Much has been written about circuit breakers, especially after their opening during the 2010 flash crash event, when they clearly did make a necessary intervention (Kirilenko et al., 2017). However, there is also the evidence of Goldstein (2015), who shows that common knowledge of what triggers a circuit breaker creates resistance around the threshold. It is debatable whether this phenomenon is good or bad.

5.1 Learning Points

Chapter two illustrates that in general noise in financial markets in not sufficient to cause strategy formation to be unsound by showing strategies are trembling hand stable. However HF traders are still able to submit orders and cancel them with the aim of disrupting the prompt functioning of the markets.

LP1.1. A first-line deterrent against spoofing strategies or attempts at momentum ignition or aggressive scalping may take the form of a minimum order resting time. However, this requirement would not assist when market volatility begins. In such instances, the circuit breakers either deter extreme price movements or halt trading in order to allow traders to update their information set. Neither minimum resting times nor circuit breakers work against one another; rather, they complement each other, as they address different issues. This learning point is particularly relevant to trading venues.

LP1.2. Exchanges that allow short selling and compound short-selling order types may wish to consider a circuit breaker that, when open, prevents the use of various short-selling strategies only. When the circuit breaker opens upon the detection of a significant downwards price movement, traders would no longer be able to bet against the value of the assets. This mechanism may be a compromise between the desire to stabilise asset prices and the desire to keep markets open. Such a device would require extensive consultation before implementation on individual trading venues. Note, this would not be possible in over the counter markets.

From a regulator's perspective it is typically the case, across the variety of structures surveyed, that a firm that develops a trading algorithm is responsible for its validation. This is partly logical, as firms possess skilled labour. However, in smaller firms, it may be that the team that develops the algorithm is also required to certify it is fit for use, creating a principal-agent problem. In addition, there is no guarantee that larger and better-resourced firms with a more technically minded compliance function would be able to avoid the principal-agent problem, as these firms may also experience the 'too-big-to-manage' problem set out in Cerpa Vielma et al. (2019).

On the one hand, trading venues would also be repositories of knowledge; however, the Haim Bodek complaint to the SEC illustrates that exchanges have a profit motive and may also struggle to fairly validate algorithms and even the compound order types they allow. Hence, although exchanges do have the necessary knowledge to self-regulate, they would often face a principal-agent problem when attempting to do so.

LP2. Self-regulatory approaches are fundamentally flawed due to the existence of a principal-agent problem. The requirements of regulators regarding the independence of internal validators could partially address this problem, although care would be needed not to disadvantage smaller firms. Furthermore, in the UK case, clearer requirements from the FCA could help to ensure an auditable minimum standard of testing. Analysis in chapter four revealed the lack of clarity in guidance produced by the PRA and FCA which could lead to firms taking a minimalist approach or trying to comply ‘on the cheap’.

National regulators also have a part to play in terms of oversight as increased monitoring activity would likely lead to increased compliance and would incentivise firms to overcome the principal-agent problem (von Wangenheim, 2004). In the UK, the FCA may request of any regulated firm the details of the algorithms it deploys in UK markets. It is a legal reequipment that a firm must respond to the request within 14 days. In Germany, the HFT Act requires algorithmic traders to register algorithms before they are deployed and provide details thereof. It is also required that firms should record the identity of the algorithm that generated each unit of order flow.

Some mandates are made concerning the elements to be included in algorithms, such as kill-switch controls (FCA, 2018). It would also be interesting to see the position check made mandatory, as it currently is in Chinese markets (Wang et al., 2019). This would, to a modest extent, slow traders down; more importantly, it may improve situational awareness if a trading entity is generating conflicting order flows. In an event where order flows conflict, the technology likely has conflicting information and should be halted until a position can be determined.

Chapter three illustrates an ARCH based model which can be used to decompose price data around events of interest. Where regulators and third parties have access to suitable data it is possible to assess the individual contribution of trading entities to the price formation process using their working price levels

LP3.1. National conduct regulators should have access to information that may assist in an investigation. Either mandatory registration or a system of spot checks should be considered in order to encourage firms to comply with good algorithmic management practices and to create a ‘spirit of transparency’.

LP3.2. National conduct regulators may wish to consider the use of RegTech approaches to increase their ability to monitor activity and ensure all filings can be scrutinised promptly. This would allow monitoring to be undertaken with a good economy of scale in the long term.

Finally, this chapter considered international regulation and the scope for it. Morelli (2017) argued that international regulation would be ‘contentious and impractical’, and the ordoliberal school of thought would suggest that views of sovereignty are a difficult issue. The EU is perhaps the most likely body to implement regional provisions for HFT. Although the EU has many Directives made against financial markets, these mostly cover consumer finance products. MiFID has a provision that is too narrow to include the vast majority of HFT activity (Hemetsberger, 2006).

LP4. Within the EU, consideration should be given to standardised guidance to those developing and deploying trading algorithms, and a similar process of disclosure should be employed to enable investigations that involve cross-arbitrage across borders (but within the bloc) to be effective.

5.2 Conclusion

This thesis has sought to assess the contribution HFT makes to financial markets in terms of contribution and disruption. Based upon this an analysis of what options are available to regulators from a analytical perspective (chapter three) and a legal/regulatory perspective (in chapter four). An updated model of market microstructure, incorporating multiple trader types, illustrates with proxy noise data that rational trading strategies are trembling hand stable, that is they are still correct to play in the presence of current noise levels. Ergo, little can be said that HFT makes markets too noise for slower traders to operate.

Chapter three sets out what little is known about the operational black box of a HFT system and their internal decision-making processes which involve streaming and co-processing technology. An ARCH type model is applied to a novel process of determining an individual traders working price level. This, using a frequent dataset, is used to demonstrate that regulators are able to deconstruct events in retrospect where needed to see how individual traders contributed to the price formation process. Given the density of data, regression coefficients give an average behaviour over 10-20 minute blocks. This method illustrates that where events can be flagged a simple technique can be used to determine if an individual trading entity may warrant further attention in line with the regulatory frameworks summarised in chapter four.

Chapter four contributes to the challenge of fairness in financial markets. Existing literature focuses upon freedom from coercion as a means of fairness. Whilst this thesis does not argue against this, it does extend the idea of fairness in terms of regulatory touch. For instance, the fact that one trader is faster than another may appear unfair to the slower trader, however by following an ordo-liberal structure one can argue that this is subjective and subject to two traders having made different choices. Rather the ordo-liberal based approach considers how negative spillovers from non-cooperative interactions can be minimised without targeting a particular trader type. This is reflected in the policy recommendations set out above which give practical advice to trading venues and regulators. Some sources may go further to slow, HFT down, however such recommendations are ruled out by the ordo-liberal approach which seeks to minimise resistance from stakeholders. Ideas such as using copper wire are debunked in chapter four.

6 References

Alexander, S. (1961) ‘Price movements in speculative markets: trends or random walks’, *Industrial Management Review*, vol. 2, no. 2, pp. 7–27.

Allen, F. and Morris, S. (1998) ‘Finance applications of game theory’, Cowles Foundation Discussion Papers, 1195 [Online]. Available at http://cowles.yale.edu/sites/default/files/files/pub/d11/d1195.pdf

Angel, J. J. (2014) ‘When finance meets physics: the impact of the speed of light on financial markets and their regulation’, *Financial Review*, vol. 49, no. 2, pp. 271–281 [Online] Available at doi: 10.1111/fire.12035

Angel, J. J. and McCabe, D. (2013) ‘Fairness in financial markets: the case of high-frequency trading’, *Journal of Business Ethics*, vol. 112, no. 4, pp. 585–595.

Aquilina, M. and Ysusi, C. (2019) *FCA Insight: Under the spotlight: high-frequency trading* [Online]. Available at https://www.fca.org.uk/insight/under-spotlight-high-frequency-trading

Arner, D., Barberis, J., and Buckley, R. (2016) ‘FinTech, RegTech, and the reconceptualization of financial regulation’, *Northwestern Journal of International Law and Business* [Preprint], vol. 35 [Online] Available at https://ssrn.com/abstract=2847806

Arnuk, S. and Saluzzi, J. (2009) ‘Latency arbitrage: the real power behind predatory high-frequency trading’, Themis Trading LLC White Paper.

Ba Fin [Germany] (2013) *Hochfrequenzhandelsgesetz* (T1, 23).

Bank of International Settlements (2021) History of the Basel Committee [Online] Available at https://www.bis.org/bcbs/history.htm

Banks, E. (2009) *Dark pools: off-exchange liquidity in an era of high frequency, program, and algorithmic trading*, 2nd edn, London, Palgrave Macmillan.

Barclay, M. J. and Hendershott, T. (2003) ‘Price discovery and trading after hours’, *Review of Financial Studies*, vol. 16, no. 4, pp. 1041–1073 [Online]. Available at doi:10.1093/rfs/hhg030

Bereby-Meyer, Y. and Roth, A. E. (2006) ‘The speed of learning in noisy games: partial reinforcement and the sustainability of cooperation’, *American Economic Review*, vol. 96, no. 4, pp. 1029–1042 [Online]. Available at doi: 10.1257/aer.96.4.1029

Black, F. (1986) ‘Noise’, *The Journal of Finance*, vol. 41, no. 3, pp. 528–543 [Online]. Available at doi:10.1111/j.1540-6261.1986.tb04513.x

Bodek, H. (2013) *The Problem of HFT*. CreateSpace.

Bodek, H. and Shaw, M. (2012) *Introduction to HFT Scalping Strategies*. Decimus Capital Markets.

Bonefeld, W. (2012) ‘Freedom and the strong state: on German ordoliberalism’, *New Political Economy*, vol. 17, no. 5, pp. 633–656 [Online]. Available at doi:10.1080/13563467.2012.656082

Bossche, F. A. M. V. den (2011) ‘Fitting state space models with EViews’, *Journal of Statistical Software*, vol. 41, no. 8 [Online]. Avaliable at doi: 10.18637/jss.v041.i08

Bouvert, A. et al. (2014) *High-frequency trading activity in EU equity markets.* European Securities and Markets Authority [Online]. Available at https://www.esma.europa.eu/sites/default/files/library/2015/11/esma20141_-_hft_activity_in_eu_equity_markets.pdf.

Box, G. E. P. and Tiao, G. C. (1975) ‘Intervention analysis with applications to economic and environmental problems’, *Journal of the American Statistical Association*, vol. 70, no. 349, p. 70 [Online]. Available at doi: 10.2307/2285379

Brown, M. S., Pelosi, M. J., and Dirska, H. (2013) ‘Dynamic-radius species-conserving genetic algorithm for the financial forecasting of Dow Jones index stocks’, in Perner, P. (ed) *Machine Learning and Data Mining in Pattern Recognition*, Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 27–41 [Online]. Available at doi: 10.1007/978-3-642-39712-7_3

Butler, T. and O’Brien, L. (2019) ‘Understanding RegTech for digital regulatory compliance’, in Lynn, T. et al. (eds) *Disrupting Finance,* Cham, Springer

International Publishing, pp. 85–102 [Online]. Avaliable at doi: 10.1007/978-3-030-02330-0_6

Cable, V. (2014) ‘Observations on the UK banking industry’, *International Review of Financial Analysis*, vol. 36, pp. 84–86 [Online]. Avaliable at doi:10.1016/j.irfa.2014.11.011

Cerpa Vielma, N. et al. (2019) ‘Too big to manage: US megabanks’ competition by innovation and the microfoundations of financialization’, *Cambridge Journal of Economics*, vol. 43, no. 4, pp. 1103–1121 [Online]. Available at doi:10.1093/cje/bez027

Chande, T. S. (2001) *Beyond technical analysis: how to develop and implement a winning trading system,* 2nd edn, New York, Wiley (Wiley Trading).

Chlistalla, M. (2011) ‘High-frequency trading: better than its reputation?’, *Deutsche Bank Research* [Online]. Available at www.dbresearch.com

Cho, D.D. et al. (2003) ‘The magnet effect of price limits: evidence from high-frequency data on Taiwan Stock Exchange’, *Journal of Empirical Finance*, vol. 10, nos. 1–2, pp. 133–168 [Online]. Available at doi:10.1016/S0927-5398(02)00024-5

Chung, K. H. and Lee, A. J. (2016) ‘High-frequency trading: review of the literature and regulatory initiatives around the world’, *Asia-Pacific Journal of Financial Studies*, vol. 45, no. 1, pp. 7–33 [Online]. Available at doi:10.1111/ajfs.12120

Code of Federal Regulations 17 § 242.612: Minimum pricing increment (n.d.). Legal Information Institute, Cornell Law School [Online] Available at https://www.law.cornell.edu/cfr/text/17/242.612 (Accessed 24 April 2019).

Cohen, D. (2012) *System and method for low latency market data* [Online]. Available at https://patents.google.com/patent/US8130758B2/en

Collins, H. (1984) *Marxism and law.* Oxford and New York, Oxford University Press.

Consolidated Tape Association (2020). *Consolidated Tape Association* [Online]. Available at https://www.ctaplan.com/index

Cruickshank, D. (2000) *Competition in UK banking: a report to the Chancellor of the Exchequer*. HM Treasury UK [Online]. Available at https://webarchive.nationalarchives.gov.uk/20050301221631/http://www.hm-treasury.gov.uk/documents/financial_services/banking/bankreview/fin_bank_reviewfinal.cfm (Accessed 12 November 2020).

Currie, W. L. and Lagoarde-Segot, T. (2017) 'Financialization and information technology: themes, issues, and critical debates – Part I', *Journal of Information Technology*, vol. 32, no. 3, pp. 211–217 [Online]. Available at doi:10.1057/s41265-017-0044-8

D'Avino, C. (2017) 'Banking regulation and the changing geography of off-balance sheet activities', *Economics Letters*, vol. 157, pp. 155–158 [Online]. Available at doi:10.1016/j.econlet.2017.05.035

DeBedts, F. (1964) *The new deal's SEC.* New York, Columbia University Press.

Dolgopolov, S. (2014) 'High-frequency trading, order types, and the evolution of the securities market structure: one whistleblower's consequences for securities regulation', *University of Illinois Journal of Law, Technology, and Policy*, vol. 2014, pp. 145–175.

Easley, D. et al. (1996) 'Liquidity, information, and infrequently traded stocks', *The Journal of Finance*, vol. 51, no. 4, pp. 1405–1436 [Online]. Available at doi: 10.1111/j.1540-6261.1996.tb04074.x

Easley, D., López de Prado, M. M., and O'Hara, M. (2011) 'The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading', *The Journal of Portfolio Management*, vol. 37, no. 2, pp. 118–128 [Online]. Available at doi: 10.3905/jpm.2011.37.2.118

Ekinci, C. and Ersan, O. (2018) 'A new approach for detecting high-frequency trading from order and trade data', *Finance Research Letters*, vol. 24, pp. 313–320 [Online]. Available at doi: 10.1016/j.frl.2017.09.020

Ersan, O. and Ekinci, C. (2016) 'Algorithmic and high-frequency trading in Borsa Istanbul', *Borsa Istanbul Review*, vol. 16, no. 4, pp. 233–248 [Online]. Available at doi: 10.1016/j.bir.2016.09.005

EUREX (2013, 5 August) *High-Frequency Trading – a Discussion of Relevant Issues* [Online]. Available at https://www.eurexchange.com/exchange-en/technology/high-frequency_trading

EUREX Deutschland (2021) *Order-to-Trade Ratio* [Online]. Available at https://www.eurex.com/resource/blob/245816/0290660651ed504641c44702d4aa3186/data/concept-order-to-trade-ratio.pdf

European Securities and Markets Authority (2020). *Short Selling* [Online] Available at https://www.esma.europa.eu/regulation/trading/short-selling

Fama, E. and Malkiel, B. (1970) 'Efficient capital markets: a review of theory and empirical work', *The Journal of Finance*, vol. 25, no. 2, pp. 383–417 [Online]. Available at doi: https://doi.org/10.1111/j.1540-6261.1970.tb00518.x|

Financial Conduct Authority (2018) *Algorithmic Trading Compliance in Wholesale Markets*, Financial Conduct Authority.

Financial Conduct Authority (2018) *Algorithmic Trading Compliance in Wholesale Markets.*

Financial Conduct Authority (2020) *Short Selling* [Online]. Available at https://www.fca.org.uk/markets/short-selling

Financial Conduct Authority (2016) *Supervision* [Online] Available at www.fac.org.uk/about/supervison (Accessed 1 March 2022).

Fischer, T. and Krauss, C. (2018) Deep learning with long short-term memory networks for financial market predictions*, European Journal of Operational Research,* vol. 270, no. 2, pp. 654-669

Foucault, T., Hombert, J., and Roşu, I. (2016) 'News trading and speed: news trading and speed', *The Journal of Finance*, vol. 71, no. 1, pp. 335–382 [Online]. Available at doi: 10.1111/jofi.12302

Foucault, T., Kozhan, R., and Tham, W. W. (2017) 'Toxic arbitrage', *The Review of Financial Studies*, vol. 30, no. 4, pp. 1053–1094 [Online]. Available at doi:10.1093/rfs/hhw103

Fouskas, V. and Gökay, B. (2012) *The Fall of the US Empire: Global Fault-Lines and the Shifting Imperial Order,* London, Pluto Press.

Fouskas, V. K. and Roy-Mukherjee, S. (2019) 'Neo-liberalism and ordoliberalism – one or two critiques? An introduction', *Critical Sociology,* vol. 45, nos. 7–8, pp. 953–965 [Online]. Available at doi:10.1177/0896920519835008

FTX Exchange (2020) *Maker vs Taker Fees* [Online] Available at https://help.ftx.com/hc/en-us/articles/360033394492-Maker-vs-Taker-fees (accessed: 17 November 2020)

German Federal Financial Supervisory Authority (2020) Algorithmic Trading and High-Frequency Trading [Online]. Available at https://www.bafin.de/EN/Aufsicht/BoersenMaerkte/Hochfrequenzhandel/high_frequency_trading_node_en.html

Goldstein, M. A. (2015) 'Circuit breakers, trading collars, and volatility transmission across markets: evidence from NYSE Rule 80A', *Financial Review*, vol. 50, no. 3, pp. 459–479 [Online]. Available at doi:10.1111/fire.12074

Greenwald, B. C. and Stein, J. C. (1991) 'Transactional risk, market crashes, and the role of circuit breakers', *The Journal of Business*, vol. 64, no. 4, p. 443 [Online]. Available at doi:10.1086/296547

Grullon, G., Michenaud, S., and Weston, J. P. (2015) 'The real effects of short-selling constraints', *The Review of Financial Studies*, vol. 28, no. 6, pp. 1737–1767 [Online]. Available at doi:10.1093/rfs/hhv013

Hagströmer, B. and Nordén, L. (2013) 'The diversity of high-frequency traders', *Journal of Financial Markets*, vol. 16, no. 4, pp. 741–770 [Online]. Available at doi: 10.1016/j.finmar.2013.05.009

Harris, L. (2013) 'What to do about high-frequency trading', *Financial Analysts Journal*, vol. 69, no. 2, pp. 6–9 [Online]. Available at doi:10.2469/faj.v69.n2.6

Hartigan, J. and Wong, M. (1979) 'Algorithm AS 136: a K-means clustering algorithm', *Journal of the Royal Statistical Society*, vol. 28, no. 1, pp. 100–108 [Online]. Available at doi: 10.2307/2346830

Haykin, S. S. (1996) *Adaptive Filter Theory*, 3rd edn, Upper Saddle River, New Jersey, Prentice-Hall International.

Hemetsberger, W. (2006) *European Banking and Financial Services Law*. Alphen on the Rhine, Netherlands and Frederick, Maryland, Kluwer Law International in association with EAPB ; sold and distributed in North, Central and South America by Aspen Pub.

Hicks, J. R. (1962) 'Liquidity', *The Economic Journal*, vol. 72, no. 288, p. 787 [Online]. Available at doi: 10.2307/2228351

HM Treasury (2019) *Financial Services Future Regulatory Framework Review* [Online]. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819025/Future_Regulatory_Framework_Review_Call_for_Evidence.pdf

Houston, J.F., Lin, C., and Ma, Y. (2012) 'Regulatory arbitrage and international bank flows', *The Journal of Finance*, vol. 67, no. 5, pp. 1845–1895 [Online]. Available at doi:10.1111/j.1540-6261.2012.01774.x

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65-91.

Johannsen, K. (2017) 'Toxic arbitrage and price discovery', SSRN Electronic Journal. doi: 10.2139/ssrn.2950161.

Jopson, B., Meyer, G., and Arnold, M. (2018) 'US regulator fines European banks for "spoofing"', *Financial Times* [Online]. Available at https://www.ft.com/content/d361a25a-0502-11e8-9650-9c0ad2d7c5b5

Kagel, J. H., Kim, C., and Moser, D. (1996) 'Fairness in ultimatum games with asymmetric information and asymmetric payoffs', *Games and Economic Behavior,* vol. 13, no. 1, pp. 100–110 [Online]. Available at doi: 10.1006/game.1996.0026

Kavassalis, P. et al. (2018) 'An innovative RegTech approach to financial risk monitoring and supervisory reporting', *The Journal of Risk Finance*, vol. 19, no. 1, pp. 39–55 [Online]. Available at doi:10.1108/JRF-07-2017-0111

Keller, A. (2012) ‘Robocops: regulating high-frequency trading after the flash crash of 2010’, *Ohio State Law Journal*, vol. 73, no. 6.

Kemp, A. G., & Reid, G. C. (1971). The random walk hypothesis and the recent behaviour of equity prices in Britain. *Economica*, *38*(149), 28-51.

Kerwer, D. (2005) ‘Rules that many use: standards and global regulation’, *Governance*, vol. 18, no. 4, pp. 611–632 [Online]. Available at doi:10.1111/j.1468-0491.2005.00294.x

Kirilenko, A. et al. (2017) ‘The flash crash: high-frequency trading in an electronic market: the flash crash’, *The Journal of Finance*, vol. 72, no. 3, pp. 967–998 [Online]. Available at doi:10.1111/jofi.12498

Kokot, S. (2004) *The Economics of Sequential Trade Models,* Berlin, Springer.

Kress, J. (2019) ‘Solving banking’s “too-big-to-manage” problem’, *Minnesota Law Review*, vol. 171 [Online]. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3348593

Lang, A. et al. (2017) *European Union (Withdrawal) Bill*. House of Commons Library [Online]. Available at https://commonslibrary.parliament.uk/research-briefings/cbp-8079

Laughlin, G., Aguirre, A., and Grundfest, J. (2014) ‘Information transmission between financial markets in Chicago and New York’, *Financial Review*, vol. 49, no. 2, pp. 283–312 [Online]. Available at doi: 10.1111/fire.12036

Lee, R. (2011) *Running the World’s Markets: The Governance of Financial Infrastructure.* Princeton, New Jersey, Princeton University Press.

Lewis, M. (2015) *Flash Boys: A Wall Street Revolt*. New York and London, W.W. Norton & Company.

Lo, A. W., Mamaysky, H., and Wang, J. (2000) ‘Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation’, *The Journal of Finance*, vol. 55, no. 4, pp. 1705–1765 [Online]. Available at doi: 10.1111/0022-1082.00265

Lui, K. and Chong, T. (2013) ‘Do technical analysts outperform novice traders? Experimental evidence’, *Economics Bulletin*, vol. 33, no. 4, pp. 3080–3087 [Online]. Available at http://www.accessecon.com/Pubs/EB/2013/Volume33/EB-13-V33-I4-P287.pdf

Macchiarulo, A. (2018) ‘Predicting and beating the stock market with machine learning and technical analysis’, *Journal of Internet Banking and Commerce*, vol. 23, no. 1, pp. 1–21.

Machiavelli, N. *The Prince*, trans. G. Bull (2003) London and New York, Penguin Books.

Magidson, J. and Vermunt, T. (2002) ‘Latent class models for clustering: a comparison with K-means’, *Canadian Journal of Marketing Research*, vol. 20, pp. 37–44.

Mahmoodzadeh, S. and Gençay, R. (2017) ‘Human vs. high-frequency traders, penny jumping, and tick size’, *Journal of Banking and Finance*, vol. 85, pp. 69–82 [Online]. Available at doi: 10.1016/j.jbankfin.2017.08.015

Manahov, V. (2016) ‘A note on the relationship between high-frequency trading and latency arbitrage’, *International Review of Financial Analysis*, vol. 47, pp. 281–296 [Online]. Available at doi: 10.1016/j.irfa.2016.06.014

Menkhoff, L. and Schmeling, M. (2010) ‘Whose trades convey information? Evidence from a cross-section of traders’, *Journal of Financial Markets*, vol. 13, no. 1, pp. 101–128 [Online]. Available at doi: 10.1016/j.finmar.2009.08.001

Moglen, E. (2020) *Chandelor v Lopus* [Online]. Available at http://moglen.law.columbia.edu/twiki/bin/view/EngLegalHist/LopusChandler

Moody’s Analytics (2011) *Regulation Guide: An Introduction* [Online]. Available at https://www.moodysanalytics.com/-/media/whitepaper/2011/11-01-03-regulation-guide-introduction.pdf

Moore, M. J. and Payne, R. (2011) ‘On the sources of private information in FX markets’, *Journal of Banking & Finance*, vol. 35, no. 5, pp. 1250–1262 [Online]. Available at doi: 10.1016/j.jbankfin.2010.10.013

Morelli, M. (2017) ‘Implementing high-frequency trading regulation: a critical analysis of current reforms’, *Business and Entrepreneurial Law Review,* vol. 6, no. 2, pp. 201–229.

Morris, G., Thomas, D., and Luk, W. (2009) ‘FPGA-accelerated low-latency market data feed processing’, *Proceedings of the 17th IEEE Symposium on High Performance Interconnects* [Online]. Available at doi: 10.1109/HOTI.2009.17

Mourad, R. et al. (2013) ‘A survey on latent tree models and applications’, *Journal of Artificial Intelligence Research*, vol. 47, pp. 157–203 [Online]. Available at doi: 10.1613/jair.3879

Narayan, P. K., Mishra, S., and Narayan, S. (2015) ‘New empirical evidence on the bid-ask spread’, *Applied Economics*, vol. 47, no. 42, pp. 4484–4500 [Online]. Available at doi: 10.1080/00036846.2015.1031870

Nasdaq (2019) *Order-to-Trade Ratio Calculation Methodology* [Online]. Available at https://cns.omxgroup.com › cds › DisclosureAttachmentServlet (Accessed 15 November 2019).

Ng, H. R. and Lam, K. (2006) ‘How does sample size affect GARCH models?’ doi: 10.2991/jcis.2006.139

Niederhoffer, V. and Osborne, M. F. M. (1966) ‘Market making and reversal on the stock exchange’, *Journal of the American Statistical Association*, vol. 61, no. 316, pp. 897–916 [Online]. Available at doi: 10.1080/01621459.1966.10482183

O’Hara, M. (2014) ‘High-frequency trading and its impact on markets’, *Financial Analysts Journal*, vol. 70, no. 3, pp. 18–27.

O’Hara, M. (2015) ‘High-frequency market microstructure’, *Journal of Financial Economics*, vol. 116, no. 2, pp. 257–270 [Online]. Available at doi:10.1016/j.jfineco.2015.01.003

O’Loughlin, D. (2013) ‘FCA: ‘Buyer beware’ is hard to defend’, *FT Adviser* [Online]. Available at https://www.ftadviser.com/2013/04/10/regulation/regulators/fca-buyer-beware-is-hard-to-defend-OrQpV8KEt9zL4Zslo8VPPP/article.html

Ormerod, C. (1939) *Dow Theory Applied to the London Stock Exchange*. London, Pitman.

Osipovich, A. (2019) ‘More exchanges add “speed bumps”, defying high-frequency traders’, *Wall Street Journal*, 29 July [Online]. Available at https://www.wsj.com/articles/more-exchanges-add-speed-bumps-defying-high-frequency-traders-11564401611

Pang, B., Lee, L., and Vaithyanathan, S. (2002) ‘Thumbs up? Sentiment classification using machine learning techniques’, *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing – EMNLP ’02*, Association for Computational Linguistics, pp. 79–86 [Online]. Available at doi: 10.3115/1118693.1118704

Patterson, S. (2012) *Dark Pools*. New York, Random House.

Peters, M. (2016) *HFT Act: Changes to the Order-to-Trade Ratio (OTR) 046/16,* Eurex Deutschland [Online]. Available at https://www.eurexchange.com/resource/blob/190210/c062933e957624476b94e0b1c2c17dcf/data/er16046e.pdf

Petrella, G. (2006) ‘Option bid-ask spread and scalping risk: evidence from a covered warrants market’, *Journal of Futures Markets*, vol. 26, no. 9, pp. 843–867 [Online]. Available at doi: 10.1002/fut.20216

Pope, H. (1910) ‘The English common law in the United States’, *Harvard Law Review*, vol. 24, no. 1, pp. 6–30.

Prudential Regulation Authority (2018) *PS12/18 Algorithmic Trading*, Prudential Regulation Authority.

Quinn Mills, D. (2002) *‘Let the Buyer Beware’ Doesn’t Protect Investors,* Harvard Business School [Online]. Available at https://hbswk.hbs.edu/item/let-the-buyer-beware-doesnt-protect-investors

Roberts, H. V. (1959) ‘Stock-market “patterns” and financial analysis: methodological suggestions’, *The Journal of Finance*, vol. 14, no. 1, p. 1 [Online]. Available at doi: 10.2307/2976094

Rummel, O. (2015) *Estimating the Implicit Inflation Target of the South African Reserve Bank Using State-Space Models and the Kalman Filter,* Centre for Central Banking Studies [Online]. Available at https://cmi.comesa.int/wp-content/uploads/2016/03/Ole-Rummel-9-Feb-Exercise-on-state-space-models-and-KF-EMF-EAC-9-13-February-2015.pdf (Accessed 21 March 2019).

Samuelson, L. and Zhang, J. (1992) 'Evolutionary stability in asymmetric games', *Journal of Economic Theory,* vol. 57, no. 2, pp. 363–391 [Online]. Available at doi: 10.1016/0022-0531(92)90041-F

Santiago Carbó-Valverde (2017) 'The impact on digitalization on banking and financial stability', *Journal of Financial Management, Markets, and Institutions*, vol. 1, pp. 133–140 [Online]. Available at doi: 10.12831/87063

Schilder, A. et al. (2005) *Compliance and the Compliance Function in Banks*, Basel Committee on Banking Supervision [Preprint].

Securities and Exchange Commission (2015) United States of America v EDGA Exchange, Inc. [Online]. Available at https://www.sec.gov/news/pressrelease/2015-2.html

Seddon, J. J. J. M. and Currie, W. L. (2017) 'A model for unpacking big data analytics in high-frequency trading', *Journal of Business Research*, vol. 70, pp. 300–307 [Online]. Available at doi: 10.1016/j.jbusres.2016.08.003

Shefrin, H. and Statman, M. (1993) 'Ethics, fairness, and efficiency in financial markets', *Financial Analysts Journal*, vol. 49, no. 6, pp. 21–29 [Online]. Available at doi: 10.2469/faj.v49.n6.21

Sheridan, I. (2017) 'MiFID II in the context of financial technology and regulatory technology', *Capital Markets Law Journal*, vol. 12, no. 4, pp. 417–427.

Siems, M. and Schnyder, G. (2014) 'Ordoliberal lessons for economic stability: different kinds of regulation, not more regulation: return of ordoliberalism – beyond the rhetoric', *Governance*, vol. 27, no. 3, pp. 377–396 [Online]. Available at doi:10.1111/gove.12046

Silber, W. L. (1984) ‘Marketmaker behavior in an auction market: an analysis of scalpers in futures markets’, *The Journal of Finance*, vol. 39, no. 4, p. 937 [Online]. Available at doi: 10.2307/2327606

Stafford, P. (2016) ‘FCA study on high-frequency trading provides few answers’, *Financial Times* [Online]. Available at https://www.ft.com/content/29286e1e-0486-11e6-a70d-4e39ac32c284

Stenfors, A. and Susai, M. (2019) ‘Liquidity withdrawal in the FX spot market: a cross-country study using high-frequency data’, *Journal of International Financial Markets, Institutions and Money,* vol. 59, pp. 36–57 [Online]. Available at doi:10.1016/j.intfin.2018.11.010

Subrahmanyam, A. (2013) ‘Algorithmic trading, the flash crash, and coordinated circuit breakers’, *Borsa Istanbul Review*, vol. 13, no. 3, pp. 4–9 [Online]. Available at doi:10.1016/j.bir.2013.10.003

Sutton, J. (1997) ‘Game theoretic models of market structure’, in Krepes, D. and Wallace, K. (eds) *Advances in Econometrics*.

The Standard (Hong Kong) (2019) ‘Goldman, Morgan Stanley seek to amend Jardine Matheson trades after stock crash’, *The Standard (Hong Kong),* 25 January [Online]. Available at https://www.thestandard.com.hk/breaking-news/section/2/120656/Goldman,-Morgan-Stanley-seek-to-amend-Jardine-Matheson-trades-after-stock-crash

UK House of Commons (2018) *EU (Withdrawal Agreement) Bill* (No. 16).

UK House of Commons (2018) *Short Selling (Amendment) (EU Exit) Regulations* (No. 1321).

Uniform Sales Act 1906 [Online]. Available at http://source.gosupra.com/docs/statute/221

Securities and Exchange Commission [US] (2005) *Regulation SHO.*

Von Neumann, J. and Morgenstern, O. (2007). *Theory of Games and Economic Behaviour*, Princeton University Press.

von Wangenheim, G. von (2004) *Games and Public Administration: The Law and Economics of Regulation and Licensing*, Cheltenham, UK and Northampton, Massachusetts, Edward Elgar.

Wah, E. and Wellman, M. P. (2013) 'Latency arbitrage, market fragmentation, and efficiency: a two-market model', *Proceedings of the 14th ACM conference on Electronic commerce – EC '13*. Philadelphia, Pennsylvania, US: ACM Press, p. 855 [Online]. Available at doi: 10.1145/2492002.2482577

Wang, J. (1994) 'A model of competitive stock trading volume', *Journal of Political Economy*, vol. 102, no. 1, pp. 127–168 [Online]. Available at doi: 10.1086/261924

Wang, S.S., Xu, K., and Zhang, H. (2019) 'A microstructure study of circuit breakers in the Chinese stock markets', *Pacific-Basin Finance Journal*, vol. 57, p. 101174 [Online]. Available at doi:10.1016/j.pacfin.2019.101174

Wang, Y. et al. (2013) 'LTC: A latent tree approach to classification', *International Journal of Approximate Reasoning*, vol. 54, no. 4, pp. 560–572 [Online]. Available at doi: 10.1016/j.ijar.2012.06.024

Wang, Z. and Zheng, W. (2015) *High-Frequency Trading and Probability Theory*. Singapore, World Scientific.

Wilson, A. (2012) *Contract Law*. Harlow, Pearson.

Wilson, R. (1985) 'Game-theoretic analyses of trading processes', Stanford, California, 474.

Woodward, M. (2017) 'The need for speed: regulatory approaches to high-frequency trading in the United States and the European Union', *Vanderbilt Journal of Transnational Law*, vol. 50, no. 5, pp. 1–44.

Yang, S. et al. (2012) 'Behaviour-based learning in identifying high-frequency trading strategies', *IEEE*, pp. 1–8 [Online] Available at doi: 10.1109/CIFEr.2012.6327783