High Frequency Trading and Herding

Servanna Mianjun Fu

A thesis submitted for the degree of Doctor of Philosophy

Essex Business School

University of Essex

April 2022

Abstract

In financial markets, the potential tendency of traders to follow some type of consensus action is referred to as herding. Traders might herd intentionally (i.e. they intend to mimic others' behaviour or follow the market consensus), or they might herd unintentionally (spurious herding). The literature shows mixed evidence of herding which mainly focuses on human traders, while herding evidence from non-human traders such as algorithmic traders and high frequency traders is absent from the herding literature. Therefore, in this thesis, the role of high frequency trading (hereafter HFT) in herding is discussed in the context of a single market and the most popular exchanges around the world.

The thesis employs quotes and trade volumes to proxy HFT in the US equity market and provide evidence that HFT induces spurious herding when trading intensity is high. Moreover, the colocation start date and HFT effective date are used from ten of the most popular global exchanges to proxy the emergence of HFT and estimate the effect of HFT on herding. Again, it is shown empirically that the emergence of HFT induces herding even during the financial crisis period. Finally, the implementation of MiFID II from the beginning of 2018 allows access to data which flags algorithmic trading under different traders. Instead of using different methods to proxy algorithmic trading, we can therefore identify each algorithmic trade and estimate the effect of different traders (i.e., human traders, algorithmic traders, and market makers) on herding. The results also demonstrate significant evidence of herding.

Overall, the thesis shows that HFT induces herding, given the increasing trading intensity. To best of my knowledge, this is the first time the herding literature has examined HFT and algorithmic trading or shown such findings.

Acknowledgements

I am extremely grateful to my supervisors Dr. Thanos Verousis and Prof. Neil Kellard for their continuous and valuable encouragement. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I would like to thank all my colleagues and faculty members at Essex Business School. It is their kind help and support that have made my study and life in the UK a wonderful time. I would also like to thank University of Essex for the studentship that allowed me to conduct this thesis. Last but not least, I would like to express my sincere gratitude to my parents, Liang Fu and Xiaodeng Han, and my brother, Wansen Fu. Without their tremendous understanding and encouragement in the past few years, this would have not been possible.

"You can never be overdressed or overeducated." – Oscar Wilde

Table of Contents

List of Figures

List of tables

Chapter 1 Thesis Introduction

Herding, broadly speaking, refers to one suppressing prior belief to follow the activities of others (intentional herding), or agents using similar strategies to analyse similar information and making the same trade (spurious herding) (Bikhchandani and Sharma, 2000). For example, a less sophisticated investor may intentionally herd to avoid the high costs of obtaining pertinent information (Chiang and Zheng, 2010). In this situation, disregarding one's own belief to mimic more sophisticated investors' behaviour or follow the market consensus could be more profitable for the less able investor. However, if many unquestioningly follow the consensus, this may lead to asset prices deviating considerably from their true fundamental values. Moreover, herding can also increase correlations among assets, and therefore reduce diversification benefits. Despite these possible individual and market-wide effects, in the extant literature, most work focuses on human traders while any herding evidence of other traders is still typically overlooked.

In the past decade, non-human traders such as high frequency trading (HFT) represents much of the trading volume in many developed markets (Malceniece, Malcenieks and Putninš, 2019). HFT refers to the activity of trading algorithms that submit orders or cancel orders in milliseconds and react extremely fast to market updates or new information. The motivation for focusing on equity market HFT in the context of herding is due to two reasons. First, HFT uses *similar* algorithm strategies to respond to any released fundamental information, analogously to non-HFT investors who spurious herd by employing *similar* analytical methods to assess the impact of the same information. Second, since herding behaviour normally reveals itself as a short-term phenomenon, HFT adoption of extraordinary high speeds to process information and submit orders could enable the detection of such behaviour in equity markets.

This thesis examines whether herding in the equity market can be explained by the emergence of, and the increased trading activities of, computerized trading. More specifically, these issues are investigated by using proxies of HFT, colocation start date, and the identification of algorithmic trading directly from the available data. Through the examination of single equity markets (i.e., the US equity market and the Athens stock exchange) and an international sample (i.e., ten exchanges around the world), we consider micro (stock-level) factors and discuss how HFT may induce herding.

In Chapter 2, the extant literature on herding is reviewed and work on HFT from the perspective of equity markets is also covered. The first concepts discussed are unconditional and conditional herding, where 'unconditional' herding refers to herding as a general case and empirically, the evidence is limited (see e.g., Christie and Huang, 1995; Chang et al., 2000; Gleason, Mathur and Peterson, 2004), while 'conditional' herding suggests the presence of herding in equity markets (see e.g., Galariotis et al., 2015; Bernales, Verousis and Voukelatos, 2020; Andrikopoulos, Kallinterakis, Ferreira, and Verousis, 2017; Voukelatos and Verousis, 2018). In this thesis, HFT acts as the condition or context to examine the existence of herding.

In Chapter 3, the primary methodology is introduced i.e., a non-linear regression that is used throughout and proposed by Chang et al. (2000) to estimate Cross-Sectional Absolute Deviation (hereafter CSAD), and which is calibrated from a linear regression by Christie and Huang (1995). Both methods are used to estimate herding towards the market consensus, but it is essential to note that CSAD is expected to change over time even if there are no herding effects. Importantly, as will be explained in more detail later, a non-linear regression can better estimate a herding specification for the relation between CSAD of stock returns and market returns; a negative and statistically significant coefficient for the non-linear term indicating the presence of herding.

In Chapter 4, the role of HFT in herding is examined within the context of the US equity market. A few studies have explored herding by using high frequency 'intraday' data (e.g., Andrikopoulos et al., 2017), and reported significant evidence of herding. This finding is in line with the notion proposed by Froot, Scharfstein and Stein (1992), who indicate that herding usually displays as a short-term phenomenon. In contrast, the extant herding literature finds very little evidence of herding in the US equity market employing low frequency data. Therefore, this chapter proposes that herding might be present at an even higher frequency in the US equity market, but such a possibility has not yet been analysed. To explain further, note that Malceniece et al. (2019) clarify that HFTs adopt market making and opportunistic strategies, automatically monitoring prices and market conditions and reacting similarly to each other due to the similar signals received. Although non-HFT might also see the same signal they cannot trade as fast as HFT. Instead, HFT can apply algorithmic trading strategy to trade with very low latency, issuing repeated trades if necessary in microseconds. To re-emphasise, these strategies applied by HFT imply that such traders will analyse and respond to a signal based on similar computer-based algorithms - HFT are unlikely to copy each other intentionally, thus we propose that HFT herds spuriously. This leads to the first research question:

Research question 1: Does HFT activity induce spurious herding?

The empirical results in Chapter 4 commence by showing that using data at a low frequency level (i.e., aggregated on a daily level in Chapter 4) is consistent with herding literature that shows that herding is not detected in the US equity market (Christie and Huang, 1995; Chang et al., 2000; Gleason et al., 2004; Chiang and Zheng, 2010). By contrast, statistically significant evidence of herding is found when employing data at a higher frequency (i.e., 5-minute and 10-minute intervals). Moreover, herding is additionally broken down into HFT intentional herding and HFT spurious herding, where intentional herding is driven by non-fundamental information and spurious herding is driven by fundamental information. Specifically, using the S&P 100 stocks to proxy HFT through number of quotes and trading volumes, we follow Galariotis, Rong and Spyrou (2015) and use Fama French return factors to separate the fundamental information effect and the non-fundamental information effect and investigate whether is herding intentionally or unintentionally induced by HFT. The results show significant evidence of unintentional (i.e. spurious) herding when HFT increases trading intensity, and the results are robust for both HFT proxies. While the extant literature provides little evidence of herding behaviour in the US equity market, these new findings suggest that HFT as non-human traders can induce herding. Additionally, it should be noted that the high trading intensity of HFT can trigger this spurious herding behaviour. Arguably, the increased intensity of HFT provides an information inefficiency, leading to the rationale for herding.

In Chapter 5, we investigate whether the emergence of HFT will induce herding using an international sample that includes ten exchanges in nine countries. Chapter 4 suggests that herding exists among HFT in the US equity market. Of course, compared with human traders, HFT has the advantage of very fast speed, especially when HFT places servers as close as possible to exchanges' infrastructure, which is called "colocation". HFT firms can acquire a colocation service from exchanges; alternatively, they might establish themselves at locations proximate to exchanges before a colocation service officially offered by them (Aitken, Cumming and Zhan, 2017). Both the colocation start date and the HFT effective date can be used to proxy the emergence of HFT and will increase HFT activities.[1] In order to look for the best price and lowest latency to act faster than their competitors, HFT firms need to locate their servers geographically close to exchanges to avoid microseconds latency. Despite this, the date on which each exchange offers colocation services is different. As we mentioned in Chapter 4, HFT applies similar strategies and thus induces herding. Therefore, we propose that increasing HFT activities can explain herding behaviour on exchanges.

Research question 2: Does the introduction of colocation provide the rationale for herding?

[1] The colocation start date refers to the date that exchanges officially offer a colocation service. Aitken et al. (2015) estimate the HFT effective date by using trade size. They identify the effect date of HFT as when four continuous months' trading size decline or the biggest single drop from the previous month on an exchange.

There are two proxies used to estimate the emergence of HFT, which is the HFT effective date and the colocation start date (i.e. the date of the colocation service officially introduced by an exchange). The HFT effective date is always earlier than the colocation start date, because HFT located trading facilities close to exchanges earlier than a colocation service officially offered by them. We follow Boehmer, Fong and Wu (2020) for our second proxy (colocation start date). These two proxies are used to generate two dummy variables and estimate the effect of HFT emergence on herding employing the full sample period, before the HFT effective date, between the HFT effective date and the colocation start date, after the colocation start date, and during the financial crisis.

The empirical evidence in Chapter 5 shows significant herding evidence across the international sample due to the emergence of HFT. Strikingly, herding is absent before HFT became effective. However, herding is detected under positive extreme market returns after HFT became effective, while herding is also present under the negative extreme market returns after colocation service is made available on exchanges. During the financial crisis, herding is only present when either an HFT effective date or the colocation start date exist during the same period. These findings suggest that the emergence of HFT induces herding - in particular, HFT starts to affect the market (in terms of herding) before exchanges introduce a colocation service, but the impact of HFT is greater after exchanges officially offer such a service.

The availability of an efficient platform for implementing algorithmic trading strategies is a precondition for HFT, which one can view as an extension of algorithmic trading that can respond to changes in news and market conditions within milliseconds.

In Chapter 6, herding is examined for the most traded stocks on the Athens Stock Exchange after the implementation of MiFID II. To the best of my knowledge, this is the first time in the literature that herding on non-algorithmic trading and algorithmic trading is investigated side-by-side. This is carried out by identifying three types of traders (i.e., human traders, algorithmic traders, and market makers). As indicated in previous chapters, algorithmic traders likely apply similar algorithm strategies when receiving common signals after macroeconomic announcements (Chordia, Green and Kottimukkalur, 2018). After the implementation of MiFID II, trading information became more transparent. According to Choi and Skiba (2015), herding is more likely to occur under higher levels of information transparency, given herding is potentially driven by agreement around similar fundamental information. This leads to the third research question:

Research question 3: Does algorithmic trading induce herding?

The novel dataset allows us to flag the algorithmic trades and, more specifically, whether the trade execution is a buy order or a sell order. After assessing trading intensity at the daily level, dummy variables are employed to indicate high trading intensity for different traders and to estimate whether the high trading intensity of algorithmic trading can induce herding. Furthermore, all listed stocks are sorted from our full sample according to market capitalization on each 31st December and investigate the size effect (small stocks and large stocks) on herding. The empirical results indicate that algorithmic traders tend to herd extreme capitalization stocks, especially when they intensively execute trades. We further decompose trades into buy-side and sell-side for each trader, and herding evidence is consistent. However, we do not detect any herding evidence for non-algorithmic traders.

This thesis makes several contributions to the extant literature by using different approaches to evidence that HFT and/or algorithmic trading induces herding at different exchanges. Among other things, the thesis contributes in the following manner:

1) We contribute on the literature by providing herding evidence in the US equity market. Generally, there is no herding evidence until we use HFT as a condition and conclude that HFT without human bias can induce spurious herding in the US equity market.

2) By following Aitken et al. (2015) and Boehmer et al. (2020) to use HFT effective date and colocation start date to proxy the emergence of HFT, we further evidence that the emergence of HFT induce herding across ten exchanges in nine countries.

3) By investigating the impact of new regulation (i.e. MiFID II) on algorithmic traders by clearly identifying algorithmic and non-algorithmic trades. This is the first occasion in the herding literature that strong herding evidence is detected for algorithmic traders.

The remainder of the thesis is organised as follows: In Chapter 2, there is a review of herding and HFT literature. In Chapter 3, there is a presentation of the main methodology employed to estimate herding. In Chapter 4, the data, estimation methodology, and empirical results are presented for the first research topic, “High Frequency Trading and Stock Herding”. In Chapter 5, the data, estimation methodology, and empirical results are presented for the second research topic, “Colocation and Stock Herding”. In Chapter 6, we outline the data, estimation methodology, and empirical results regarding to the third research topic, “Herding on Different Traders”. Finally, in Chapter 7, the thesis is concluded.

Chapter 2 Literature Review

In this section, we first demonstrate the background of herding behaviour. Second, some literature examines herding as a general case without considering any conditions in the equity market and finds little evidence about herding. Therefore, some literature fills this gap by involving different conditions and shows herding presence in the equity market. We discuss the empirical results for both unconditional herding and conditional herding in different subsections. Eventually, our last subsection of literature review is to discuss HFT.

2.1 Herding Background

Conceptually, herding behaviour appears in financial market when investors discard their private signals to resort or mimic other's trading behaviour following interactive observation of their activities (Hirshleifer and Teoh, 2003). There are many reasons that will induce herding. Banerjee (1992) and Bikhchandani et al (1992) argue if investors herd on other's information, meanwhile suppressing their own private signals will prevent from the information being incorporated in the public information pool. This results in the market price shaped by limited information and very likely to present information cascades (i.e. herding behaviour caused by imperfect information). This is often driven by the anticipation of informational payoffs (Devenow and Welch, 1996).

It is also possible that herding is motivated by reputation and compensation concerns. Analysts or managers might follow others who have higher abilities and better performance (Froot et al., 1992; Hirshleifer, Subrahmanyam and Titman, 1994). In this case, analysts or managers infer information from others' previous excellent recommendations, meanwhile, neglect their own information in order to achieve a better reputation and compensation (Scharfstein and Stein, 1990; Trueman, 1994). In addition, De Bondt and Teoh (1997) propose relative homogeneity among fund managers such as their educational background and professional framework can lead to correlated trading activities, considering propensity of fund managers have willingness to follow diversified investment styles (e.g. momentum / contrarian; value / growth etc) (see also, Bennett, Sias and Starks, 2003).

Generally, there is a substantial herding empirical literature focuses on two main strands in equity markets (Spyrou, 2013).[2] On the one hand, research examine herding of institutional investors (Bennett et al., 2003; Gavriilidis, Kallinterakis and Ferreira, 2013). On the other hand, aggregate market data has been used in empirical studies to investigate herding towards the market consensus (Chang et al., 2000; Chiang and Zheng, 2010).

[2] This is the comprehensive review paper of the herding literature. Also referencing two more recent studies from Andrikopoulos et al. be(2017) and Frijns and Huynh (2018). In addition to the equity markets, herding effects also have been examined in other empirical studies from different markets, for example, mutual funds (Lakonishok, Shleifer and Vishny, 1992; Wermers, 1999; Sias, 2004; Shyu and Sun, 2010), options (Bernales et al., 2020; Voukelatos and Verousis, 2019), bonds (Bikhchandani and Sharma, 2001; Galariotis, Krokida and Spyrou, 2016; Cai et al., 2018), and Exchange Traded Funds (Gleason et al., 2004).

2.2 General Unconditional Herding Empirical Results

Unconditional herding refers to the research examines where investors might consistently cluster around the market consensus, while conditional herding investigates whether herding is more likely to present during specific states of the market. Previous studies of herding mainly focus on equity markets and found mixed results. Most studies of the herding literature examine the cross-sectional dispersion of equity returns to measure whether the return of individual stocks tend to cluster around the market consensus. Christie and Huang (1995) firstly indicate that investors rely on overall market conditions for their investment decision making process. Rational asset pricing models predict that the absolute value of the market return of cross-sectional dispersion returns will increase during normal periods, as traders will trade diversely based on their own private information (Christie and Huang, 1995). Individuals tend to suppress information to imitate collective behaviour in the market, when the market is experienced extreme movements. Under these conditions, individual stock returns prone to cluster around the overall market return.

Following the approach proposed by Christie and Huang (1995), number of studies have found no evidence of significant herding in the US equity market (Chang et al., 2000; Gleason et al., 2004; Chiang and Zheng, 2010). Also, Gleason et al. (2004) examine Exchange Traded Funds (ETFs) in the US and find no evidence to support ETF investors to herd under extreme market conditions.

However, Gavriilidis, Kallinterakis and Montone (2021) employ a similar methodology and conclude significant herding behaviour in the US equity markets, when the US presidents are unpopular and their policies are controversial. Although Chang et al. (2000) reject herding behaviour in the US and Hong Kong stock market, instead, they conclude traders tend to herd around the aggregate market consensus in emerging markets such as South Korea, Taiwan, and (to a lesser extent) Japan. In contrast, Zhou and Lai (2009) focus on Hong Kong stock market and find evidence of herding behaviour in small stocks while investors have propensity to sell stocks rather than to buy stocks. Regarding to Chinese market (Shanghai and Shenzhen stock exchanges), Demirer and Kutan (2006) study whether investors making investment decisions by following the market consensus or relying on private information when market goes down. They reports that there is no herding behaviour in the Chinese market, which suggests investors in Chinese stock markets may rationally make investment decisions. However, Tan et al. (2008) argue herding behaviour arising in the Chinese stock market during periods of both up and down market conditions, especially display in A-share investors.

2.3 Conditional Herding Empirical Results

As herding as a general case (i.e. unconditional herding) absence in the equity market on previous empirical studies, Chiang and Zheng (2010) and Andrikopoulos et al. (2017) evidence the investment decisions of foreign investors in the equity market depend on international market conditions. In order to explore whether herding is more pronounced during certain periods, the market consensus of the US and the UK leading stocks have been examined by Galariotis et al. (2015). They employ cross-sectional stock returns dispersion to examine herding effects and conclude investors tend to herd when macroeconomic information release and financial crisis arise. During early stage of financial crisis, they conclude the US equity market has herding spill-over effects to the UK equity market.

Likewise, Bernales et al. (2020) support their hypothesis of conditional herding during a period of market stress in option market, where cross-sectional dispersion is significantly lower than the expected index option return. Similarly, Voukelatos and Verousis (2018) propose to utilise extracted information from the option market to interpret the conditional herding behaviour in the US equity market. They reject herding as a general case, as unconditional herding failed to explain whether traders prone to herd or not when pricing individual stocks. Instead, the authors find evidence of conditional herding where stock return dispersion is significantly lower, compared to expected market return. In this case, although herding under extreme market condition is not pronounced, they emphasise the herding behaviour of investors is more closely clustered around the market consensus when relatively pessimistic view has been displayed in the option market's trading activities.

2.4 Literature Review of HFT

Given new and high technology access to equity market, it allows traders to process information and submit orders at extraordinary high-speed. The holding periods can be measured by milliseconds or even microseconds, leading to large trading volumes and algorithmic strategy become major force in equity market, so-called HFT. Securities and Exchange Commission (SEC) refers HFT as "professional trading with proprietary ability to employ strategies to generate numerous trading on daily basis". Such HFT firms rely on high-speed advantage and sophisticated computer programs to generate, route and execute orders. Technology substantially changed the competition nature on trading venues. After trading transfer from human to algorithm trading by machine, searching costs become almost trivial (e.g. the quote spread is severely reduced). A fast machine allows the market-making strategy to quickly update quotes when public information arrived, thus reduces adverse selected risk. HFT might be an important role to link multiple exchange markets and makes the real competition between markets become possible (Stoll, 2001).

In the following sections, we review the relevant literature results of HFT. HFT is implementing algorithm strategy by using computers. Therefore, research findings of algorithm trading are relevant to the effects of HFT. Researchers generate different predictions based on their assumptions to examine how HFT has important effect on market quality.

2.4.1 Positive Effects of HFT on Market Liquidity

Some empirical results prone to report positive effect on market liquidity. Hendershott, Jones and Menkveld (2011) first reveal the relationship between algorithm trading and HFT. They use the introduction of auto-quote at NYSE in 2003 as an instrument for algorithm trading and find algorithm trading have positive effect on liquidity as well as faster price discovery. Malceniece et al. (2019) extend the similar method, using the electronic messages normalized by trading volume as a proxy of algorithm trading. They indicate the significant increases on co-movement in returns and in liquidity is driven by HFT. In addition, positive effect of algorithm trading on liquidity has been found in Deutsche Boerse stocks (Hendershott and Riordan, 2013) and in a global sample (Boehmer et al., 2020).

2.4.2 Positive Effects of HFT on Informational Efficiency

Moreover, HFT also has positive effects on informational efficiency. Through the analysis of dataset of NASDAQ-identified HFT between 2008 and 2010, Carrion (2013) and Brogaard, Hendershott and Riordan (2014) implement different methods and conclude the same results. They indicate the days with higher HFT intensity are related with higher informational efficiency. In order to facilitate price discovery, liquidity demanding orders will be submitted by HFT towards to two directions (i.e. permanent price changes and temporary pricing errors). Moreover, Chaboud, Chiquoine, Hjalmarsson and Vega (2014) find algorithm trading improves the frequency of arbitrage opportunities and the autocorrelation of high frequency returns, which are the two measures of informational efficiency.

2.4.3 Negative Effects of HFT

In contrast, other studies demonstrate the negative effects of HFT. Boehmer, Li and Saar (2015) indicate HFT follows trading strategies against transitory price pressures to further liquidity supply. The result shows the extent of HFT competition has negative effect on short-term volatility. Contrary to the traditional view of limit orders providing liquidity to the market, Hasbrouck and Saar (2009) argue such quality or useful liquidity provided by HFT is very short-lived since many limit orders are fleeting out in electronic markets. Dichev, Huang and Zhou (2014) show HFT can result in undesirable levels of volatility. In contrast, Boehmer et al. (2020) document liquidity is reducing, and volatility is worsening when there is greater algorithm trading intensity. As a result, algorithm trading becomes less beneficial when market making is difficult. Hendershott and Menkveld (2014) and Brogaard et al. (2014) using the same method to separate the permanent and temporary components in price changes and conclude that HFT does not responsible to temporary shocks in the market. Imbalance of non-HFT trading orders can lead to extreme price movements, and HFT can be able to stabilize prices (Brogaard et al., 2018). By analysing the flash crash on 6th May 2010, Kirilenko, Kyle, Samadi and Tuzun (2017) argue HFT makes the flash crash worsen but it did not cause the crash.

Chapter 3 Framework to Detect Herding

3.1 Basic Herding Specification

Christie and Huang (1995) propose a common measure of herding behaviour by using Cross-Sectional Standard Deviation (CSSD) of returns. They indicate overall market conditions will affect the investment decision-making process of investors. Thus, they believe herding behaviour will be more prevalent under extreme market movements, and individual stock return prone to cluster around the market consensus under the market stress with relatively low dispersion. A similar approach is proposed by Chang, Cheng and Khorana (2000), who argue the linear and growth relation between dispersion and market returns might disappear and making increase or decrease of non-linearly become possible, if traders prone to follow aggregate market behaviour in the period of large average price changes. To this end, they use a non-linear regression specification (i.e. Cross-Sectional Absolute Deviation) to calibrate the model of Christie and Huang (1995).

We follow Chang et al. (2000)'s method to estimate herding towards to the market consensus in the equity market based on the cross-sectional dispersion of stock returns around the market return. The relation between the Cross-Sectional Absolute Deviation (CSAD) of stock return and market return is expressed as:

$$CSAD_{t/m} = \frac{1}{N}\sum_{i=1}^{N} |R_{i,t/m} - R_{mkt,t/m}| \tag{3.1}$$

where N is the number of stocks i included in the cross-section at day t or at minute m, $R_{i,t/m}$ and $R_{mkt,t/m}$ is the stocks return and market return for each time interval day t or minute m, respectively. We compute the theoretical basic herding specification of stock return from Eq. (3.1) on each five-minute, ten-minute, and one-day frequency for the whole sample, where

CSAD is regressed only against market return. CSAD as the measure of dispersion, quantified the average proximity of stock returns from the market consensus and displayed heterogeneity on the aggregate market level. However, CSAD will expect to change over time even if there are no herding effects. Thus, measurement of herding effects in the equity market depends on conditional level of cross-sectional dispersion. As Chang et al. (2000) identified, the magnitude of CSAD should be directly related to the magnitude of contemporaneous market returns.

As a starting point in the analysis, Chang et al. (2000) suggest there must have positive relation between cross-sectional dispersion of stock return and the market return under the moderate assumptions of the Capital Asset Pricing Model (CAPM). The expected stock return under CAPM can be expressed as follows:

$$\mathrm{E}[r_{i,t}] = r_{f,t} + \beta_i \times E\left[r_{mkt,t} - r_t\right] \tag{3.2}$$

where $r_{f,t}$ is the return on the zero-beta asset (the risk-free rate) and β_i is the stock's time-invariant systematic risk at time *t*. Then, let β_{mkt} refer to the systematic risk of an equally weighted market portfolio, hence $\beta_{mkt} = {}^{1}/_{N} \sum_{i=1}^{N} \beta_i$. So, the absolute deviation of stock *i*'s expected return at time *t* from the average portfolio return can be expressed as:

$$|r_{i,t} - r_{mkt,t}| = |\beta_i - \beta_{mkt}| E[r_{mkt,t} - r_{f,t}] \tag{3.3}$$

Thus, we can define the expected cross-sectional absolute deviation of stock *i* returns (ECSAD) at time *t* as follows:

$$ECSAD_t = \frac{1}{N} \sum_{i=1}^{N} |\beta_i - \beta_{mkt}| E\left[r_{mkt,t} - r_{f,t}\right] \tag{3.4}$$

Dispersion and the time-varying market expected return has increasing linear relationship, which can be easily written as:

$$\frac{\partial ECSAD_t}{\partial E[R_{mkt,t}]} = \frac{1}{N}\sum_{i=1}^{N}|\beta_i - \beta_{mkt}| > 0 \tag{3.5}$$

$$\frac{\partial^2 ECSAD_t}{\partial E[R_{mkt,t}]^2} = 0 \tag{3.6}$$

Based on the above results in Eq. (3.5) and Eq. (3.6), additional regression parameter is required to test for herding effects, to capture any possible non-linear relation between stock return dispersion and the market return in the equity market. CSAD and $R_{mkt,t}$ have been used to proxy for the unobservable ECSAD and $E[R_{mkt,t}]$. The conditional version of the CAPM only to construct the presence of a linear relationship between ECSAD and $E[R_{mkt,t}]$, instead, we detect the presence of herding behaviour through the average relation between realized CSAD and $R_{mkt,t}$. The positive (negative) linear relationship between CSAD and $R_{mkt,t}$ implies that we expect higher (lower) market return (in absolute terms) will have higher (lower) dispersion. If large price movements during a period are associated with a decrease (or less than proportional increase) in the stocks' dispersion around the market consensus, then we identify traders are more likely to herd.

Then, a non-linear (OLS) regression is estimated as given in Eq. (3.7):

$$CSAD_{t/m} = \beta_0 + \beta_1\left|R_{mkt,t/m}\right| + \beta_2 R^2_{mkt,t/m} + \varepsilon_{t/m} \tag{3.7}$$

where CSAD is the absolute value of each stocks' deviation related to the equally-weighted average return of market portfolio, and $|R_{mkt,t/m}|$ is the absolute value of an equally weighted realized return of all available stocks on day t or at minute m.

Following Eq. (3.7) to test unconditional herding, we then augment the equation accordingly to examine each of the research question in Section 4.5, Section 5.3, and Section 6.4.

3.2 Positive and Negative Values of the Herding Coefficient

Under rational pricing assumptions, β_1 is expected to be positive and β_2 would be insignificant. The positive (negative) linear relationship between CSAD and $R_{m,t}$ implies that we expect higher (lower) market return (in absolute terms) will have higher (lower) dispersion. This is due to the individual assets differ in their sensitivity to the market return. Therefore, if there is no significant difference between market returns and cross-sectional dispersion, it implies rational pricing and no herding. On the other hand, in the presence of herding, security returns will not deviate too far from the overall market return. More specifically, herding would present itself in statistically significant negative values of β_1 (strong herding) and β_2 (moderate herding), i.e., the return dispersion is too low due to suppression of investors views and investors simply following the market consensus. This implies market-wide herding.

However, if the value of β_2 is statistically significantly positive, it indicates the case of excessively high cross-sectional return dispersion and implies 'negative herding' (Christie and Huang, 1995; Chiang and Zheng, 2010; Gebka and Wohar, 2013). This implies return dispersion is higher than the prediction of rational pricing model during period of large price movements. Thus, it does not indicate that traders to suppress their individual views and will not follow the market consensus. Unlike the market-wide herding, they seem to do the opposite to largely ignore information conveyed by the market-wide price movements. Instead, they focus on dominant views from subset of traders in an excessive and exaggerated way. This behaviour meets the following situations: localised herding, retreat from the market during market stress, and overconfidence (Gebka and Wohar, 2013).

Chapter 4 High Frequency Trading and Stock Herding

4.1 Introduction

Given most of the previous papers failed to detect significant herding in the US equity market by considering herding as an unconditional general behaviour, while a small part of empirical papers introduces other contemporaneously observed conditions from equity market to detect herding effects (Galariotis et al., 2015; Voukelatos and Verousis 2019). [3] We therefore motivate and focus on HFT as a condition to estimate herding clustering around the market consensus. More specifically, we investigate whether HFT activities induce herding in the US equity market. In addition, if HFT herds significant, it explains intentional herding or spurious herding.

We contribute significantly to the ongoing debate on the empirical herding literature by incorporating the HFT proxies with herding behaviour in the US equity market. We investigate the time period of 2015 to 2017 in the US equity market by using S&P 100 stocks which represents the concentration of HFT activities. We use millisecond frequency dataset to construct five-minute, ten-minute, and one-day frequencies dataset. This allows us to estimate the influence of HFT on equity market based on different intensities. We address the gap in the literature by providing evidence at the first time to reveal that herding is increasing in the US equity market, conditioning by HFT intensities. To the best of our knowledge, this is the first time that HFT characteristics have been examined in the herding literature. We concentrate on the connection between herding and HFT in the equity market and try to understand whether HFT can be an interpretation of presence of conditional herding in the US equity market.

Our main finding is that increased HFT activity induces spurious herding in the US equity market. Regarding to the results of basic herding specification of five-minute, ten-minute, and one-day frequencies, we did not detect any herding evidence from daily frequency.

[3] Galariotis et al. (2015) documented that the US investors tend to herd conditioned on the market releasing important macroeconomic information. While Voukelatos and Verousis (2019) examine conditional herding by using information extracted from the options market to evidence significant herding in the US stock market.

However, herding behaviour is significant at both five-minute and ten-minute frequencies. Our one-day frequency result is consistent with previous empirical findings that suggest the absence of unconditional herding behaviour as a general case in the US equity market. This rejection of herding indicates the cross-sectional dispersion of stock returns is increased with the magnitude of market returns, which is in line with the theoretical predictions by Chang et al. (2000). However, we find strong evidence to support the hypothesis that HFT induces herding behaviour on five-minute and ten-minute frequencies. In addition, our further results also evidence HFT intensities can explain herding on stock-day in the US equity market. These findings reveal the relationship between cross-sectional dispersion and the market returns which cannot be explained by the initial theory.

Besides these findings on conditional herding, we also present some evidence of a strong herding when we take extreme values of HFT proxies. We create daily subsamples when we let HFT proxies take extreme values at lower 5% tail and upper 5% tail of its distribution. In contrast to theoretical predictions and our findings on the daily full sample, we report that dispersion is no longer positively related to market returns. Theoretically, larger absolute market returns are expected to be associated with higher dispersion of individual stock returns if herding behaviour absence. However, on days with extremely low HFT volume, extremely low and high trade size, larger market movements are actually associated with lower dispersion. This indicates the stock returns cluster more closely around the market consensus.

Our results have implications for HFT companies and markets. On the one hand, the effects of herding triggered by HFT can be utilized by HFT companies for the purpose of formulating style strategies. On the other hand, our results render that herding behaviour not only presence among individual investors, but also in HFT since the increased HFT activity results in inefficient information to the market. These results are also of key interest to the market regulators, since exchange markets can allow the transmission of HFT incidents across each other with potentially destabilizing effects.

The rest of this thesis is organized as follows. In Section 4.2, we review the theoretical and empirical literature for intentional herding and spurious herding, respectively. In Section 4.3, we present the reasonable connection between herding behaviour and HFT, as well as our research questions and hypotheses. In Section 4.4, we discuss our data and define the construction of HFT proxies we adopt. In Section 4.5, we demonstrate our empirical results, while Section 4.6 discusses the empirical results. The final section concludes.

4.2 Literature review[4]

Herding behaviour exists for different reasons, thus Bikhchandani and Sharma (2001) classify herding as intentional and/or unintentional (spurious). In this section, we discuss the classifications and the empirical results caused by these two types of herding behaviour. In order to overview the theoretical and empirical framework regarding herding behaviour of investors from the equity market, Bikhchandani and Sharma (2001) distinguish herding behaviour as *intentional herding* (i.e. investors intentionally copy each other's actions) and *spurious herding* (i.e. investors face similar information sets to make a decision driven by fundamental information).

4.2.1 What is Intentional Herding and Relevant Empirical Results

Herding is intentional while the choice is driven by a potential positive externality element (a benefit), and usually presupposes the opposing view from one to another. If a manager chooses to herd in order to reap informational payoffs or career / reputational payoffs it is considered as intentional herding (Scharfstein and Stein, 1990; Devenow and Welch, 1996; Trueman, 1994; Clement and Tse, 2005). For example, if an investor realises his information is invaluable or his ability to process information is inadequate compared to other investors, then he will consider himself to be in an asymmetry situation compared to other investors. In order to free-ride the information superiority of other investors and extract information payoff, he will rationally copy other's trading activities (Devenow and Welch, 1996). But if he chooses to discard his private signals and prefers to follow others, this will slow down the signal-flow to the market (information blockage), presents a poor public pool of information, and leads to the evolution of information cascades (Banerjee, 1992; Bikhchandani, Hirshleifer and Welch, 1992).

[4] See Chapter 2 for a comprehensive literature review.

Regarding career / reputational payoffs, a manager who is not confident in his skills has every interest to mimic his peers' trade, since this will allow him to cover his inability during the period of evaluation performance (Scharfstein and Stein, 1990). Adding to the above, Economou, Gavriilidis and Kallinterakis (2015) pay attention to frontier market and find herding evidence of fund managers who expect to receive informational payoffs and / or professional payoffs. Holmes, Kallinterakis and Ferreira (2013) examine herding behaviour in Portugal and conclude it is intentional due to reputational concerns and information cascades. They afford a reasonable explanation, which indicate fund managers from small exchanges in Portugal are more likely to trust "good" managers who have better behaviour and strengths from developed countries (e.g. US and UK). Similarly, Gavriilidis et al. (2013) identify herding is intentional in the Spanish market, manifesting on fund managers who herd for information and career concerns.

4.2.2 What is Spurious Herding and Relevant Empirical Results

In contrast, spurious herding implies that investors exhibit convergence behaviour due to relative homogeneity and characteristics trading. Investors can display correlation in their trading activities because of relative homogeneity, which refer to a similar education background or professional qualification (De Bondt and Teh, 1997), use similar indicators in their analyses (Froot et al., 1992; Hirshleifer et al., 1994), and the common regulatory framework that they obey (Voronkova and Bohl, 2005; Olivares, 2008). These exhibit convergence on investment decisions through analysing information signals or indexes on the same way, resulting in correlated trades as a result of commonality. As such, leading to a similar response from traders rather than simply imitation (Gavriilidis et al., 2013; Holmes et al., 2013).

Characteristic trading is another reason for spurious herding, referring to any strategies (such as contrarian and momentum strategies) on specific equity characteristics (such as past performance and size). In some cases, if investors pursue a contrarian strategy, it exemplifies that one will expect the correlation in other's trading activities, whereby one will herd going long on recent winners and short on recent losers (Guney et al., 2017). According to the above, these are not intentional imitative herding behaviours due to the same trading style, which render the herding behaviour is unintentional among investors (Bennett et al., 2003). [5]

[5] See literature review for HFT in Chapter 2.

4.3 Propose Potential Herding Behaviour in HFT

4.3.1 Background of Research Idea

The bulk of the herding literature mainly analyse data with a frequency ranging from daily to yearly (i.e. low frequency trading), but recent studies (e.g. Andrikopoulos et al., 2017) have explored the scope of herding research basis of high frequency data (e.g. intraday herding). This development is consistent with the notion of Froot et al. (1992), who indicate that herding usually display as a short-term phenomenon. The development of technology and regulation have opened the door for multi-traders to use computers for trading competition, which allow high-speed quote update. Herding research attempts to capture intraday trading dynamics. Under this circumstance, very little evidence of herding by using intraday data has been found in the advanced market (Gleason et al., 2004, Andrikopoulos et al., 2017). Nonetheless, significant intraday herding has been found among smaller stocks in advanced market during market downturns (Zhou and Lai, 2009), and in emerging markets (Blasco, Corredor and Ferreruela, 2011, 2012). In addition, Andrikopoulos et al. (2017) first evidence cross-border markets have versatile herding dynamics on intraday level.

Although the above is expected to include high frequency data on intraday level to detect herding effects, HFT proxies have not been considered as the instruments in such research. As a fast trading technology, HFT is not a recent phenomenon in the equity market. Pagnotta and Philippon (2012) demonstrate exchanges agree to invest in fast trading technology, as non-HFT speed is too low in equilibrium participation and outcomes are generally inefficient when all venues allow market structure and speed to arise endogenously.

4.3.2 Why HFT may Induce herding?

Instead of examining unconditional herding, this study uses HFT as a condition to examine the herding behaviour. We attempt to understand whether different intensities of HFT activity would induce herding in the US equity market, as well as the fact of whether it is intentional herding or spurious herding. Given the lack of research evidence on this issue, it is unclear whether is possible to find significant herding behaviour in HFT activity in equity market from a theoretical perspective. To this end, we now attempt to present reasonable theoretical justifications in favour and/or against the presence of herding behaviour in the HFT activity. We focus on HFT intensity responds to changes in market conditions in the context of herding is motivated by the following reasons.

The case in favour of HFT in equity market have impact on herding behaviour is founded on the following arguments. First, Malceniece et al. (2019) indicate ‘market making strategy’ as one of the HFT trading strategy. HFT market makers are better able to automatically monitor the prices and market conditions of other stocks compared to non-HFT market makers, and then exploit this information to optimally setting quotes (Hendershott and Riordan, 2013).

Second, another category of HFT trading strategy is ‘opportunistic strategy’ (e.g. momentum and comprising arbitrage) based on Malceniece et al. (2019)’s clarification. HFT can use opportunistic strategy to handle long and short positions for different stocks at the same time, and also adopt similar strategies for each other based on similar signal (Chaboud et al., 2014; Biais and Woolley, 2011; Boehmer, Li and Saar, 2018). Jarrow and Protter (2012) argue opportunistic HFT will generate destabilizing effect when it coordinates unknowingly on common signal. More specifically, they indicate common signal could be the difference between future and forward prices of a stock index, which has been wrongly believed as an arbitrage opportunity or electronic news generated by the financial press. HFT employs same information can capture the common signal which might be the realization of market related event or mispricing. Although non-high frequency traders can see this signal, they cannot trade quickly enough according to their observation since unspecified constraints (e.g. lack of financial resources). By contrast, HFT could trade immediately through construction on this signal before it incorporates into market price. It implies HFT trading strategies do not need to be predictable, instead, HFT can have option regarding to market information set because of high-speed advantage.

Finally, Jarrow and Protter (2012) indicate HFT follows algorithmic trading strategy to frequently submit ask (buy) quotes at different prices and amounts, and then watch what happens to the ask (buy) by observing the algorithm. HFT will cancel the rest of quotes to render their phantom quotes, once ask (buy) is purchased (sold). HFT repeats this in microseconds and learn the algorithmic strategy to front-run it. For instance, if HFT front-runs and knows that one will buy a stock, HFT will buy the stock first and then sell it at a higher price than was paid.

Hypothesis I: HFT activity induces herding.

As we mentioned in the previous part, there is spurious herding and intentional herding. According to the previous arguments, the following point of views and associated assumptions are corresponding to different HFT strategies. First, HFT market making strategy will respond by analysing the same data using similar methods rather than mimic other traders (Froot et al., 1992; Hirshleifer et al., 1994). This towards to the market consensus through aggregate market price and market activity data (Choi and Skiba, 2015; Cai, Han, Li and Li, 2018).

Second, based on the description of opportunistic strategy, we assume HFT will do the same trade at the same time, once HFT sees the common signal. HFT acts independently based on this common signal, unknowingly but tends to collectively behave as large traders. The stock price will decline (increase) when HFT collectively sells (buys) the stock. According to these statements above, herding behaviour might induce by HFT, but in higher intensity compared to non-HFT since HFT has a bulk of trades happened in the short time. [6] Moreover, this correlated trading activity of HFT is responding to the same trading strategy.

Hypothesis II: HFT activity induces spurious herding.

---

[6] This assumption is in line with Froot et al. (1992), who indicate herding behaviour is very likely to show at the short time phenomenon.

Finally, considering algorithmic trading strategy is done by using computers, HFT activity follows the computer routes to program. Unlike herding behaviour of non-HFT caused by human's behaviour bias, HFT herds on algorithm basis and therefore is not human related. All computers used for HFT will trade simultaneously, since the potential profit will be taken by the increasing of HFT quoting activities which program the fastest and run algorithm the fastest in order to achieve the best bid price or the best ask price. To this end, we propose HFT activity is not human related and thus would not be an intentional behaviour, because of the computer routes and algorithm designs are similar. However, the more intense HFT activities might induce herding behaviour, as the similar program and algorithm would be repeated more often to increase quoting activities. Therefore, HFT imitates others while HFT follows each other for profit purpose, which is in consistent with the notion of herding but not intentionally.

Hypothesis III: High intensity HFT activity increases spurious herding.[7]

[7] Due to the fact of HFT, statistical significance of herding is expected to increase as HFT intensity increases. We split the data into different quintiles based on the trading intensity to examine the hypothesis III.

4.4 Data and HFT Proxies

4.4.1 Data

We examine HFT herding in the US equity market for a sample period from January 2015 to December 2017 due to the restriction of the raw data availability. Our sample is limited to the 100 most-traded stocks that together constitute the S&P100, a leading stock index. We limit our sample to the S&P100 constituents because HFT tends to concentrate activities on the most-traded stocks (Boehmer et al., 2020). We access all history messages of trades and quotes are time-stamped to millisecond. Our dataset includes the following information: sample's name, trade date, trade and quote time in millisecond, activity types (i.e. trade or quote), trade and quote price, trade and quote size, trade volume. In addition, daily data of Fama-French return factors are collected from Kenneth R. French Data Library. [8]

When processing the data, we identify trade or quote activity between 09:30am and 16:00pm (ET) which are the core trading hours in the US equity market as valid data. [9] Considering there will be one hour different when Daylight Saving Time begins and ends, we postpone one hour for our selected core trading hours when Daylight Saving Time ends. [10] We also exclude any national holidays. The first 10 minutes and last 10 minutes of each trading day are dropped, in order to minimize any overnight and market closing effect. We then create three frequencies for each five-minute, ten-minute, and one-day. Moreover, we winsorizing delete all variables at 1 and 99 percentiles within our sample. [11]

We create several new variables and calculated based on information from our dataset on each five-minute, ten-minute and one-day frequency for each stock *i* at time *t* as follow: 1)

---

[8] Data for three factors (HML, SMB, and MOM) will use later for our regression.

[9] Eastern Time (ET) refers to Eastern Standard Time (EST) when Daylight Saving Time (DST) ends (winter), and refers to Eastern Daylight Time (EDT) when Daylight Saving Time begins (summer).

[10] When DST began, our core trading hours between 9:30 and 16:00 (EDT). While DST ended, core trading hours are adjusted from 10:30 to 17:30 (EST). In 2015, DST began on 8th March and ended on 1st November. In 2016, DST began on 13th March and ended on 6th November. In 2017, DST began on 12th March and ended on 5th November.

[11] We winsorize trade price and trade size, as well as quote (bid and ask) price and quote size.

If the activity type is trade, it means there are missing values for bid/ask price/size. We fill the missing values by using the lag value of bid/ask price/size; 2) We calculate the average value for trade price and trade volume, quote (bid and ask) price and quote size; 3) We calculate trade frequency, bid updates, and ask updates, the last two will only count for an update when the bid or ask price changes; 4) We estimate quote spread as $quote\ spread_{i,t} = average\ ask\ price_{i,t} - average\ bid\ price_{i,t}$;[12] 5) We create mid-quote by using two different methods: a) $mid\_quote_{i,t} = \frac{average\ bid\ price_{i,t} + average\ ask\ price_{i,t}}{2}$ or b) using last data of bid/ask price for each frequency instead of using average price; 6) In addition, we use these two types of mid-quote to estimate log return and simple percentage return for each stock.[13] Since the results are similar from these returns, therefore, we adopt log return estimating by the first method of mid-quote calculation. The returns for each stock *i* are calculated as $R_{i,t} = 100 * (\ln(midquote_t) - \ln(midquote_{t-1}))$ at time *t*; 7) Eventually, we estimate market return for each stock i as $R_{mkt,t} = \sum_{without\ stock\ i} \frac{R_t}{99}$.

4.4.2 Proxies for HFT

4.4.2.1 HFT Number of Quote Updates

We use two proxies for HFT activity. First, we follow Conrad, Wahal and Xiang (2015) to use number of quote updates as our HFT measure:

$$NQ_{i,t} = \sum bid\ update_{i,t} + ask\ update_{i,t} \tag{4.1}$$

where $NQ_{i,t}$ is number of quote updates for stock *i* on trading day *t*, which estimate by the accumulation of bid update and ask update for stock *i* on trading day *t*. We then calculate

[12] This is followed by Hasbrouck and Saar (2009) and Boehmer et al. (2020), who indicate quote spread as the time-weighted average quoted spread (ask price minus bid price).
[13] Log return equals log value of mid-quote divided by lag mid-quote, while simple percentage return equals the difference between mid-quote and lag mid-quote divided by lag mid-quote.

equally-weighted average number of quotes across all stocks on day *t*. It refers to the number of changes that arise in the best bid or offer (BBO) price or in the quoted size at these prices during a specific time interval. We construct this proxy manually for each frequency rather than rely on the official NBBO. The reason to do so is some quotes which change more than once per second, are precluded from the NBBO set under Regulation National Market System (NMS). Many exchanges have rules to ignore such quotes for trade through protection, such as NYSE Arca Rule 5210).

4.4.2.2 HFT Volume

The second proxy is similar to Hendershott et al. (2011) and Boehmer et al. (2020), $HFTvolume_{i,t}$, estimating as the negative of trading volume (in USD 100) divided by the number of quote messages (defined as quote updates):

$$HFTvolume_{i,t}, = -\frac{dvol_{i,t}}{messages_{i,t}} \tag{4.2}$$

where $dvol_{i,t}$ is consolidated trading volume in USD 100 for stock *i* on trading day *t*, and $messages_{i,t}$ is the number of quote messages. [14] We compute average trading price and trading volume at one minute interval and then aggregate all observations for stock *i* on day *t.* for the regression, we calculate equally-weighted average value across all stocks for this proxy on day *t.* Messages include total number of quotes and number of trades on day *t*. Considering orders are placed at a very high speed and algorithm continuously searches and exploits trading opportunities through computers, large number of messages will be generated in each stock-day. Therefore, increasing in this proxy implies an increase in HFT activity.

This proxy is appropriate for HFT, since it provides a continuous scale of relative HFT intensity for exchanges and can recognize differences across exchanges as an absolute measure.

[14] We use number of quote updates to represent the number of quote messages, which is our first HFT proxy.

In addition, exploiting the relative measure of HFT can identify the nature of ‘‘fast’’ or ‘‘low-latency’’ trading. HFT gains by taking advantage of fast speed compared with non-HFT in the same market, so this relative and continuous measure of HFT can well capture this contrast. This measure focuses on best quotes and trades rather than messages related to all orders.

4.5 Methodology

4.5.1 Testing Framework for Herding

We first test herding by using our intraday data at three frequencies.[15] Under the null hypothesis of no herding, the relationship between CSAD and market return should be linear, and the dispersion should be motivated only by the magnitude of market returns (i.e. $\beta_1 > 0$ and $\beta_2 < 0$). As Chang et al. (2000) suggest, the methodology would classify herding as the case where the coefficients of non-linear terms will be negative and statistically significant.

By contrast, we define the presence of strong herding as the case where the stock return dispersions expect to decrease with the magnitude of market return. Under this type of herding, traders will herd more closely around the market consensus during the period of large price movements. This alternative hypothesis of strong herding would suggest the linear coefficients of market returns taking the wrong signs (i.e. $\beta_1 < 0$ and / or $\beta_2 > 0$), reflecting the dispersion decreases with the magnitude of market return.

HFT tends to follow the market consensus or prices individual stocks independently of the market consensus and cannot be indicated by low or high CSAD levels. Based on Eq. (3.7)., we first test hypothesis I and examine whether HFT induces herding on daily interval *t* and at minute interval *m*. Then we use HFT proxies to split the data into lower 5% and upper 5% of HFT intensity, and augment Eq. (4.3) to examine herding behaviour under extreme HFT intensity as follows:

$$\begin{aligned} CSAD_{t/m} = \beta_0 &+ \beta_1(1 - D_{t/m})R_{mkt,t/m} + \beta_2 D_{t/m}R_{mkt,t/m} \\ &+ \beta_3(1 - D_{t/m})R^2{}_{mkt,t/m} + \beta_4 D_{t/m}R^2{}_{mkt,t/m} + \beta_5 D_{LOW,t/m} \\ &+ \beta_6 D_{UP,t/m} + \varepsilon_{t/m} \end{aligned} \tag{4.3}$$

[15] Five-minute, ten-minute, and one-day frequency.

where $D_{t/m}$ is a dummy variable equals one when the market return $R_{mkt,t/m}$ is negative on day t or at minute t, otherwise equals zero, $D_{LOW,t/m}$ and $D_{UP,t/m}$ are dummy variables that takes the value of one when the market return is stay in the lower 5% tail or upper 5% tail of its distribution on day t or at minute t. We will also reproduce the empirical analysis for D_{LOW} and D_{UP} at the alternative 2.5% and 1% tails of its distribution, but there is no significant difference with the results of 5% tail. These two dummy variables are useful to test for herding in extreme market conditions.

If strong herding absence under Eq. (4.3), then we define moderate herding as the case where stock return dispersion increases at a decreasing rate with an increase in the market return. However, this increase would be consistently lower than what would expect given the actual magnitude of the market consensus, indicating that traders are likely to herd when pricing individual stocks. The hypothesis of moderate herding translates to the coefficients of squared market returns being negative and statistically significant (i.e. $\beta_3 < 0$ and / or $\beta_4 < 0$).[16]

Furthermore, we define herding in extreme market conditions similar to the case of Christie and Huang (1995), where CSAD is significantly low when the stock returns are in the 5% upper and 5% lower tails of its distribution. Dispersion will be significantly low during the large price movements if traders tend to herd more closely around the market consensus under extreme market conditions. This hypothesis will be supported if the coefficients are statistically significant and negative (i.e. $\beta_5 < 0$ and / or $\beta_6 < 0$).

Finally, we define conditional herding as the case when we split CSAD for lower 5% and upper 5% based on HFT proxies. Thus, we expect stock return dispersion will be significantly lower than suggest market return. Under the moderate assumption of the CAPM, dispersion should be only driven by market return through beta. Our proxy herding variables should not have an incremental impact on CSAD. Overall, if conditional herding presence, a

[16] We follow similar definitions of strong and moderate herding from Chang et al. (2000). The definition of herding under extreme market conditions which proposed by Christie and Huang (1995) also employed in this study.

lower conditional expectation of CSAD would imply that stock returns prone to cluster around the market consensus more closely (compared to what the market returns would suggest). Significant coefficients would compose evidence for conditional herding, meaning extremely low and / or extremely high HFT activity will induce herding behaviour in the equity market.

4.5.2 Decomposition of Intentional Herding and Spurious Herding

Bikhchandani and Sharma (2000) defined *spurious herding* as when investors very likely to herd with each other to make similar decisions as they react to the same changes for *fundamental information*. In contrast, the case of investors based on *non-fundamental information* may imitate others' trading activities will define as *intentional herding.* In order to explore the propensity of HFT and to test hypothesis II and hypothesis III, we follow Galariotis et al. (2015) to decompose CSAD measure to deviation based on fundamental information and deviation due to reaction to non-fundamental information. Specifically, this can be established by introducing return factors from Fama and French (1995,1996) and Carhart (1997), who evidence critical fundamental information may affect investor decisions on market level.[17] Given that Fama French return factors are available on daily basic, we aggregate all variables on day *t*.

Hence, we run the following empirical specification:

$$CSAD_t = \beta_0 + \beta_1\left(r_{mkt,t} - r_{f,t}\right) + \beta_2 HML_t + \beta_3 SMB_t + \beta_4 MOM_t + e_t \qquad (4.4)$$

where *HML* is the High (book-to-market ratio) Minus Low value return factor, *SMB* is the Small (market capitalization) Minus Big size return factor, and *MOM* is the monthly momentum factor.

[17] This hypothesis has been supported by empirical results. For example, Liew and Vassalou (2000) examine future Gross Domestic Product (GDP) of ten international markets by adopting the High Minus Low (HML) and the Small Minus Big (SMB) factors, and conclude significant information acquired through import these factors. Also, Gregory et al. (2003) indicate the UK market has been detected a positive correlation between HML and future GDP growth. Kessler and Scherer (2010) establish a strong relation between momentum and the macro economy.

[18] We first calculating three proxies of HFT for one-day frequency and then splitting to several HFT intensities. Considering CSAD and all the fundamental information factors are included in Eq. (4.4), the residual can be reasonable identified as the cross-sectional deviation preclude fundamental information. Intuitively, the residual can be thought as a measure of clustering HFT which reacts based on non-fundamental information. We term this deviation as $CSAD_{NONFUND}$:

$$CSAD_{NONFUND,t} = e_t \tag{4.5}$$

According to the above, HFT responds to the same changes in fundamental information can be measured as the difference between the total CSAD and the $CSAD_{NONFUND}$. We denote this term as $CSAD_{FUND}$ as follow:

$$CSAD_{FUND,t} = CSAD_t - CSAD_{NONFUND,t} \tag{4.6}$$

where $CSAD_{FUND}$ is a measure of HFT responding clustering to fundamental information. Hence, we are able to distinguish fundamental information deviation ($CSAD_{FUND}$) from total CSAD and use it to proxy spurious herding. Similarly, employing the deviation due to non-fundamental information ($CSAD_{NONFUND}$) to test intentional herding. Therefore, this study defines measure of $CSAD_{FUND}$ and $CSAD_{NONFUND}$ as dependent variables to estimate two regressions similar to Eq. (4.4) as follow:

$$CSAD_{FUND,t} = \beta_0 + \beta_1 \left| R_{mkt,t} \right| + \beta_2 R_{mkt,t}^2 + \varepsilon_t \tag{4.7}$$

[18] According to Carhart (1997), MOM can be calculated by the equal weighted average of the highest performing firms minus the equal weighted average of the lowest performing firms, lagged one month. If a stock has positive average returns for its prior 12 months, then it can be classified showing momentum.

$$CSAD_{NONFUND,t} = \beta_0 + \beta_1 |R_{mkt,t}| + \beta_2 R^2_{mkt,t} + \varepsilon_t \tag{4.8}$$

In the case of herding behaviour presence, the coefficient of non-linear term from Eq. (4.7) and Eq. (4.8) should be negative and significant as we mentioned above. After decomposed CSAD as $CASD_{FUND}$ and $CSAD_{NONFUND}$, we can use these as measure of clustering due to HFT herds on fundamental information (spurious herding) or herds on non-fundamental information (intentional herding). We estimate Eq. (4.7) and Eq. (4.8) by separating dependent variables based on different HFT intensities on the whole panel, to test hypothesis III of whether increased HFT activity induces herding behaviour in the US equity market.[19]

[19] We use lower 5%, quantile 1, quantile 2, quantile 3, quantile 4, and upper 5% to imply HFT intensity from low to high for the overall daily frequency.

4.6. Empirical Evidence

4.6.1 Descriptive Statistics

Figure 4.1 (5-minute CSAD) and Figure 4.2 (10-minute CSAD) both continuously drop down from approximately 0.47 to 0.07 within core trading hours in a day. It indicates that CSAD value is incredibly high while the market open and then gradually reduced. Quote update as one of the HFT proxies (Figure 4.3) has extremely high value while there is flash crashes or market movements. But HFT volume (Figure 4.4) as another HFT proxy is relatively stable, except unexpected low volume in August 2017. [20]

[Figure 4.1 to Figure 4.4 around here]

Table 4.1 includes three panels to present the summary of statistics for cross-sectional absolute deviation and HFT proxies. Panel A shows several key statistical measures for CSAD, R_{mkt}, $|R_{mkt}|$, R^2_{mkt} of S&P100 constituents in three frequencies. Market return dispersion has smaller standard deviation from lower frequency (one-day) to higher frequency (5-minute). 5-minute interval has the smallest minimum value of CSAD out of three frequencies but has the same smallest value of market return as 10-minute interval. In addition, both higher frequencies (5-minute and 10-minute) have the same maximum value for CSAD and market return, which have almost two times CSAD than one-day frequency. Panel B outlines the statistic results for HFT proxies on one-day interval. The HFT volume distribution is negatively highly skewed than quote update. Panel C lists the values of two HFT proxies for different intensities.

[Table 4.1 around here]

[20] HFT has the first biggest flash crash on 18th March 2015, following the continue flash crash on 21st April 2015. The second biggest flash crash is happening on 24th August 2015. On 2nd November 2015, S&P closes above 2,100 for the first time since August. Another big market movement was on 20th January 2016 when Dow Jones closes down 245 points in day of drama.

4.6.2 Is Herding Significant in the US Stock Market?

Table 4.2 presents the results based on estimation from Eq. (7) for three different frequencies (i.e. five-minute interval, ten-minute interval, and one-day interval) for the whole sample period. As the table shows, the coefficients of β_1 are positive and statistically significant at 1% level for all three frequencies. This demonstrates that the cross-sectional absolute deviation of stock return increases with the magnitude of the market return. This finding contrasts with CAPM-based theoretical predictions, but it would not constitute evidence of herding. As Table 4.2 shows, β_2 appears negative and significant on 1% level for both five-minute interval and ten-minute interval, which indicates herding behaviour presence in the US equity market on intraday level. However, β_2 is negative but not significant on one day frequency. This is in line with the findings from previous literature which indicate herding does not detect on low frequency (daily) data in US equity market (Christie and Huang, 1995; Chang et al., 2000; Gleason et al., 2004; Chiang and Zheng, 2010). Based on the result above, Table 4.2 provides an affirmative evidence of intraday herding presence in the US equity market, while herding absence on daily level. We will then test herding under extreme market conditions in the next subsection.

[Table 4.2 around here]

4.6.3 Is Herding in the US Equity Market Subject to the Size Effect?

As we mentioned earlier, if there is an increasingly linear relationship between CSAD and market return, it suggests the null hypothesis I of no herding. We estimate herding behaviour under ups and downs market from Eq. (4.3). The results from Table 4.3 present some initial support for the null hypothesis I of no herding. As we can see, the coefficients of β_1 and β_2 for three frequencies are statistically significant at the 1% level. Cross-sectional absolute deviation is positively related to positive market returns and negatively related to negative market returns ($\beta_1 > 0$ and $\beta_2 < 0$). Therefore, we did not find any evidence for strong herding hypothesis.

The coefficients of interest for moderate herding are statistically and negative values of β_3 and β_4. In Table 4.3, the coefficient β_3 of squared positive market return is positive and significant on one-day frequency. This suggests cross-sectional dispersion increases at an increasing rate when the market return is positive, however, it is not in line with the theoretical predictions which suggest the presence of herding behaviour for significantly negative coefficient. This implies the return dispersion on the daily interval is higher than the prediction of rational pricing model. The values of β_4 are negative and statistically significant for all three frequencies, indicating that cross-sectional dispersion increases at a decreasing rate and herding is presence.

Finally, limited evidence of herding is found under extreme market conditions in Table 4.3. The coefficients of the tail dummies (β_5 and β_6) are positive on one-day frequency, which indicate the cross-sectional dispersion tends to be larger on days of extreme market returns. But herding is absence since the coefficient is positive and insignificant and in line with the rational pricing assumptions. In contrast, the coefficients of the tail dummies are negative and significant at 1% level on five-minute frequency, suggesting the presence of herding behaviour under extreme market conditions on five-minute intraday level. In addition, coefficient β_5 on ten-minute frequency is significantly negative. This indicates intraday herding is more likely to herd around the market consensus during extreme downward movements compared to upward movements on ten-minute level. However, we still did not find any herding evidence on daily level, and therefore we will use HFT proxies as our conditions to further detect herding in the following subsections.

[Table 4.3 around here]

4.6.4 Is Herding Significant on HFT Proxies' Extreme Values?

We now focus on daily level in order to test our hypothesis III on whether conditional herding (i.e. consider the role of HFT activity) is absent in the US equity market. We further run Eq.

(4.3) without the tail dummies of extreme market conditions, while we let HFT proxies take extreme values at lower 5% and upper 5% of its distribution to access corresponding one-day frequency data for our variables.

We present the empirical results in Table 4.4. When we estimate herding for the daily subsample of the first HFT proxy (i.e. quote update), CSAD is found to be positively related to positive market returns and negatively related to negative market returns for both low and high extreme tails. This suggests dispersion is found to increase with the magnitude of the market consensus. But the coefficients are not statistically significant, and thus not support herding presence. Regarding to the second HFT proxy (i.e. HFT volume), we find the coefficients of low HFT volume is negative for positive returns ($\beta_1 < 0$) and positive for negative returns ($\beta_2 > 0$). It suggests that CSAD decreases with the magnitude of market drop. Although coefficient β_1 is statistically insignificant, β_2 is positive and significant on lower 5% which supports the evidence for the alternative hypothesis of strong herding during periods of particularly low HFT volume. Intuitively, HFT seems to induce herding while HFT volume is on its lower 5% distribution.

[Table 4.4 around here]

4.6.5 Is HFT Intentional Herding or Spurious Herding?

We further split the daily CSAD into different intensities by using HFT proxies to test herding behaviour. Table 4.5 presents the results from Eq. (4.7), Eq. (4.8), and the total CSAD measures (see Eq. (3.7)). $CSAD_{NONFUND,t}$ is non-fundamental driven from Eq. (4.5) and the fundamental driven $CSAD_{FUND,t}$ from Eq. (4.6) under different intensities of HFT proxies. These are estimated from Eq. (4.4) that we discussed above. When the total CSAD is employed, there is no evidence for herding on all intensities before the decomposition (β_2 is not negative and significant). This is in line with our previous results of unconditional herding absence on daily level, as well as consistent with the results on previous literature (Christie and Huang, 1995; Chang et al., 2000; Gleason et al., 2004; Chiang and Zheng, 2010).

After we decompose CSAD as $CSAD_{NONFUND,t}$ and $CSAD_{FUND,t}$, we run Eq. (4.7) and Eq. (4.8) again based on different intensities split by HFT proxies (quote update and HFT volume). Interestingly, when deviations are decomposed to deviations due to non-fundamental and fundamental factors, the results indicate that HFT herds on fundamental information when both proxies fall on quantile 3 intensity (i.e. 50% < proxy < 75%). Coefficient β_2 is negative and statistically significant at 10% level for both quote update and HFT volume. These results not only suggest HFT activity induces herding, but also evidence increased HFT activity induces spurious herding which is driven by fundamental information.

Based on the results, our consideration of conditional herding by using HFT proxies to split different HFT activity intensities, as well as the distinction of herding to intentional herding and spurious herding are relevant. We would erroneously assume there is no herding exist in all cases if we do not break the impact down to its different components. For example, this is the case for quantile 3 intensity of both HFT proxies where there is spurious herding that is not picked up before the break-down. It might cause by cancelling-out or averaging effects.

[Table 4.5 around here]

Taken together, our findings suggest HFT induces herding while quote update and HFT volume is relatively high. This provide support to the concerns expressed by Brogaard et al. (2014) that increased HFT activity intensity results in higher information inefficiency in the equity market. Meanwhile, imperfect information is one of the reasons that triggers herding behaviour in the equity market (Banerjee, 1992; Bikhchandani et al., 1992). In addition, Boehmer et al. (2018) indicate HFT is better able to extract information signals from the market. Given that HFT utilizes algorithm trading strategy through similar computer routes, HFT is driven by fundamental information. This reflects spurious herding by using similar strategies and analysing similar information among HFT.

4.7 Conclusion

We investigate for the first time in the literature that HFT as a condition for inducing herding behaviour in the US equity market. Our main finding is that increased HFT activity intensity induces spurious herding in the market. We establish causality by exploiting intraday data from January 2015 to December 2017 and construct three different frequencies, following by constructing two HFT proxies as our condition to split dataset as different HFT activity intensities.

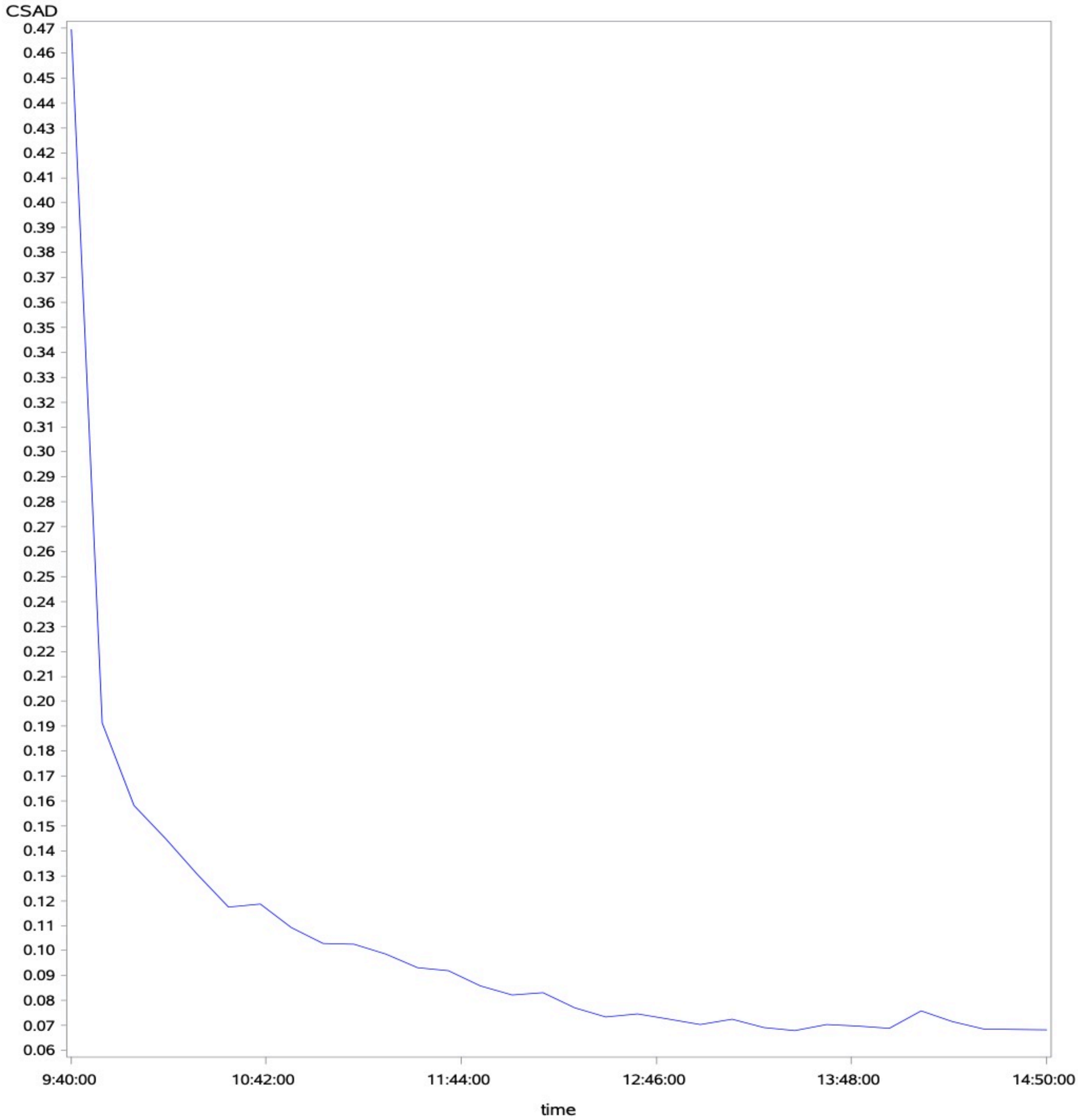
We fail to detect significant herding on one-day frequency, which is consistent with other studies. Surprisingly, herding evidence is found on 5-minute and 10-minute frequency, which might due to the higher frequencies can catch more volatility. In addition, we detect spurious herding on the 75 percentiles of HFT activity intensity for both HFT proxies. This implies HFT tends to follow each other while its quoting activity and trading volume is more intense. Our results suggest that herding behaviour is associated with increased HFT activity.

Our findings help paint a more complete picture of herding behaviour in the equity market, showing that herding behaviour is not only human-related, but also can be induced by HFT without human behaviour bias. This contributes to the literature that some results of absence herding behaviour need to be reconsider, since non-human behaviour could also be a condition to induce herding. However, because HFT has been introduced to exchange markets at different time, in order to clarify the herding effects of HFT in cross exchange markets is an important empirical question that is left for future work.

4.8 Figures and Tables

Figure 4.1

5-minute interval of CSAD

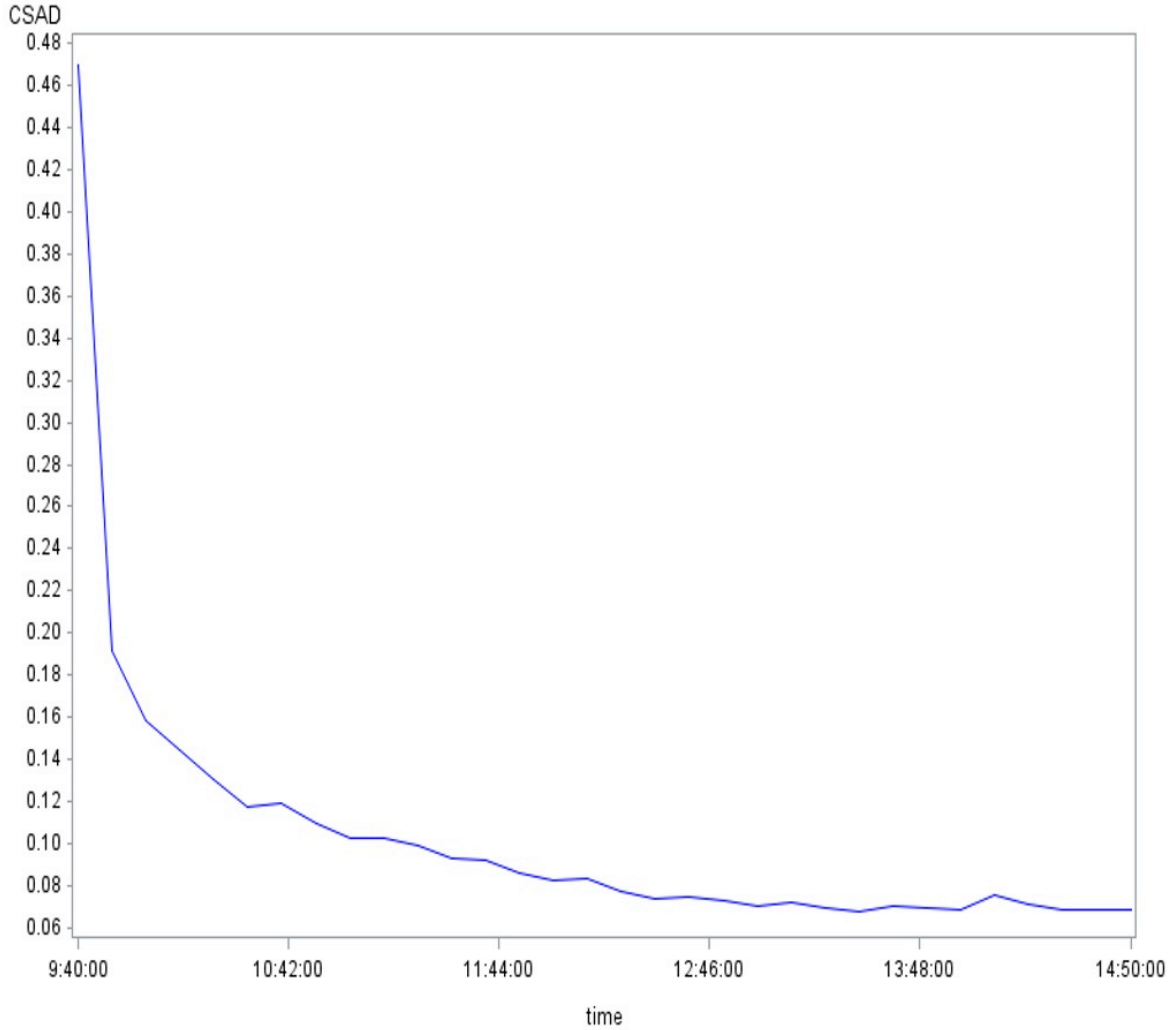


The figure plots the cross-sectional dispersion of five-minute S&P100 constituents' stock returns. The sample runs core trading hours from 9:40 to 14:50.

Figure 4.2

10-minute interval of CSAD

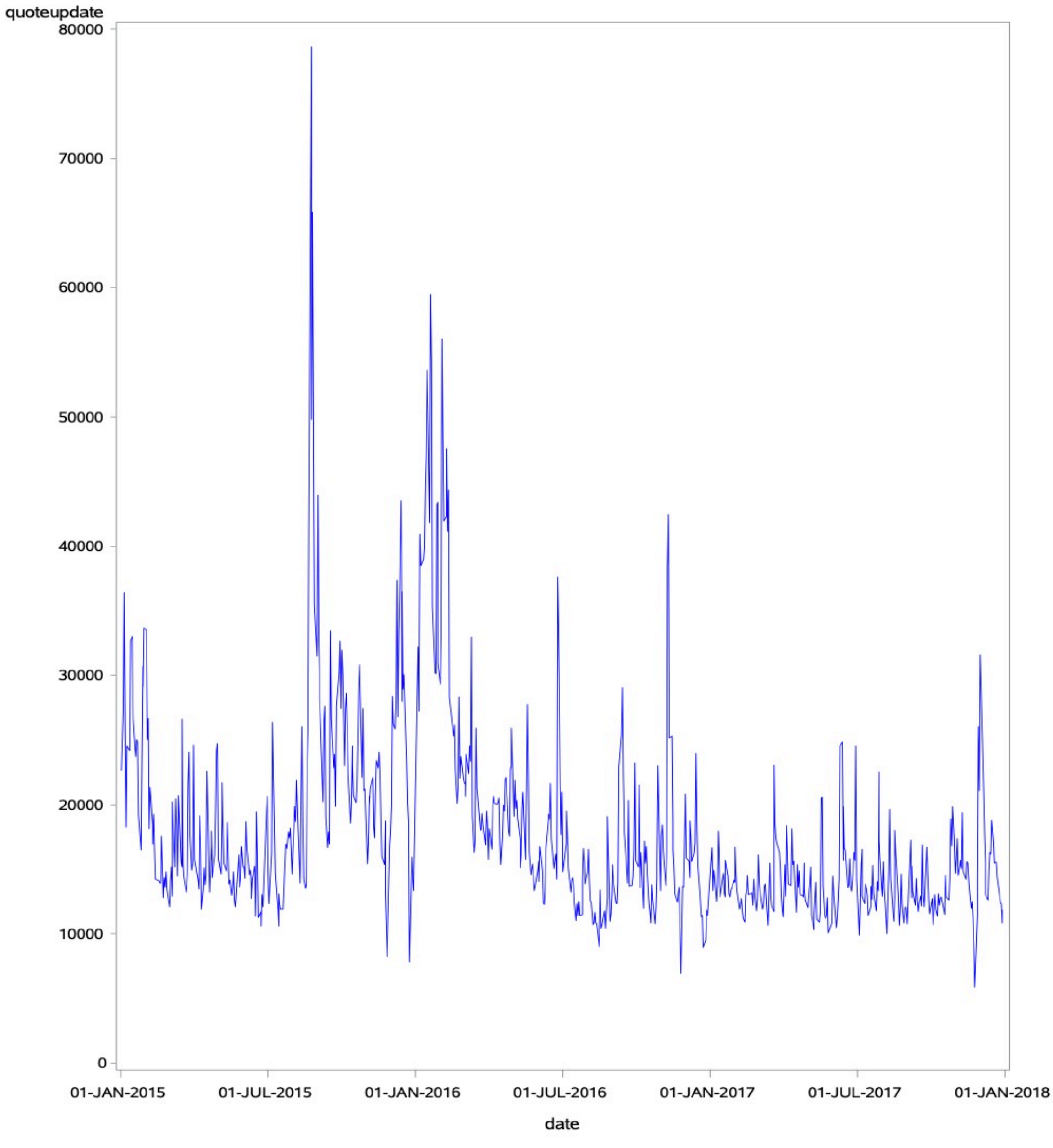
The figure plots the cross-sectional dispersion of ten-minute S&P100 constituents' stock



returns. The sample runs core trading hours from 9:40 to 14:50.

Figure 4.3

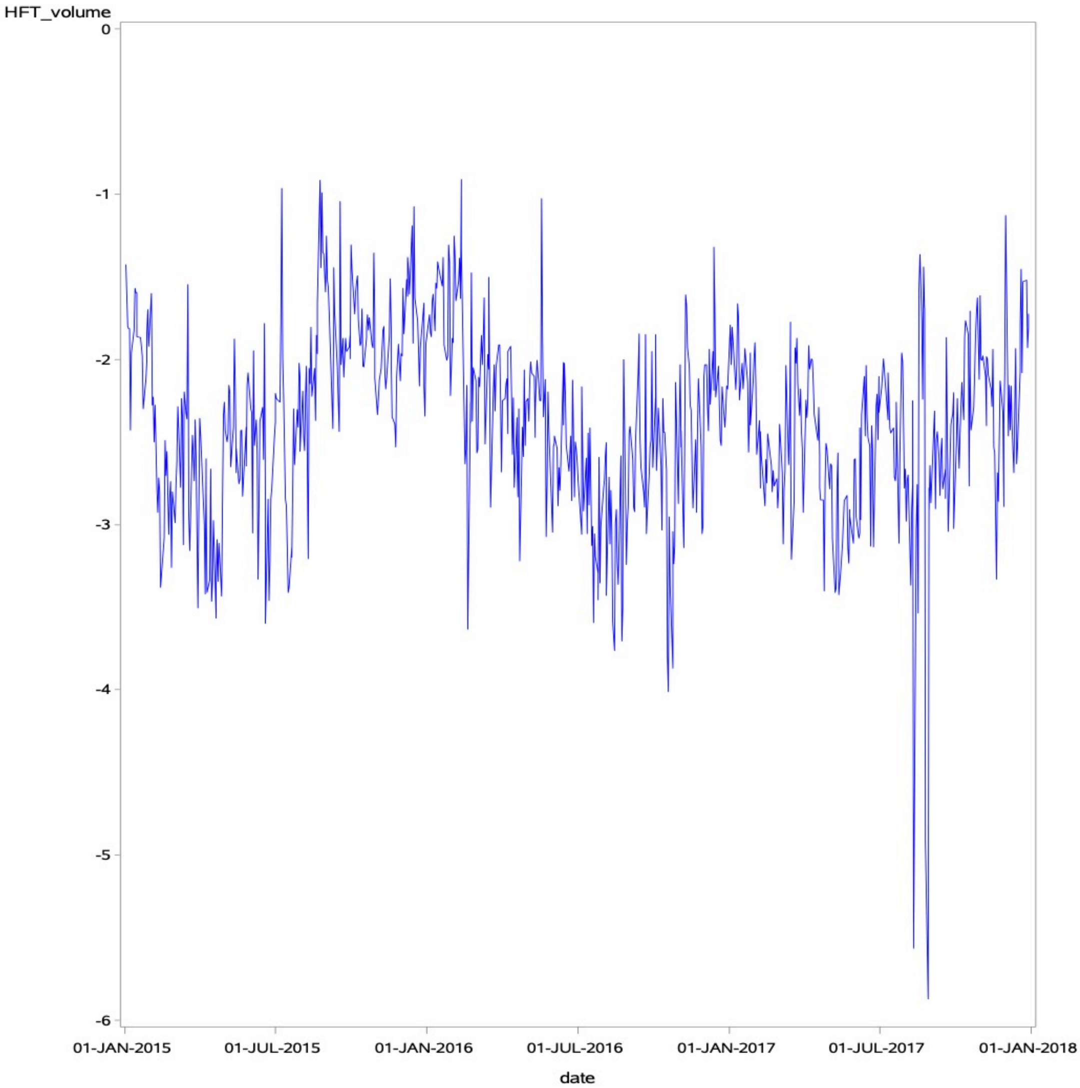
Daily plot of quote updates



The figure plots the daily distribution of quote update between 2nd January 2015 to 31st December 2017.

Figure 4.4

Daily plot of HFT_volume



The figure plots the daily distribution of HFT volume between 2nd January 2015 to 31st December 2017.

Table 4.1 Descriptive statistics

Panel A: Statistics for CSAD, R_{mkt}, $|R_{mkt}|$, R_{mkt}^2 for five-minute, ten-minute, and one-day interval

	Frequency	CSAD	R_{mkt}	$\|R_{mkt}\|$	R_{mkt}^2
Mean	5 m	0.0675	0.0002	0.0418	0.0066
	10 m	0.0981	0.0005	0.0619	0.0132
	1 d	0.5502	0.0224	0.3325	0.2355
Standard deviation	5 m	0.0649	0.0817	0.0702	0.1209
	10 m	0.0865	0.1152	0.0972	0.1704
	1 d	0.1719	0.4850	0.3537	0.6629
Minimum	5 m	0.0085	-4.8864	1.3058844E-6	1.705334E-12
	10 m	0.0168	-4.8864	2.1039955E-6	4.426797E-12
	1 d	0.2145	-3.2492	0.0003	1.0438339E-7
Maximum	5 m	3.5453	1.7082	4.8864	23.8775
	10 m	3.5453	1.7082	4.8864	23.8775
	1 d	1.8563	1.9116	3.2492	10.5574

Panel B: Statistics for daily HFT proxies of S&P100

Variable	N	Mean	Standard deviation	Min	Max	Skewness	kurtosis
Quote_update	77873	18259.63	17198.48	0	537972	4.2756	46.3153
HFT_volume	77863	-2.3945	11.1416	-1784.24	0	-58.5638	8502

Panel C: HFT proxies split into four equal-sized quantiles on one-day frequency

	Low 5%	25%	50%	75%	High 5%
Quote_update	11186.9368	13207.5208	15596.4854	20548.9904	33378.1714
HFT_volume	-3.3542	-2.7251	-2.3532	-2.0033	-1.5139

Notes: Quote_update refers to number of quote updates, HFT_volume estimates as the negative of trading volume divided by number of quote updates.

Table 4.2 Herding in the S&P 100 as a whole

	Five-minute frequency	Ten-minute frequency	One-day frequency
$\lvert R_{mkt} \rvert$	0.0065***	0.0060***	0.0028***
	(171.65)	(113.16)	(9.47)
R_{mkt}^2	-0.0003***	-0.0002***	-0.0002
	(-15.26)	(-8.32)	(-1.53)
R^2	0.444	0.420	0.259

The table presents OLS results for the full sample period between January 2015 and December 2017 for S&P 100 stocks based on the following non-linear regression: $CSAD_{t/m} = \beta_0 + \beta_1 \lvert R_{mkt,t/m} \rvert + \beta_2 R_{mkt,t/m}^2 + \varepsilon_{t/m}$, where CSAD is the Cross Sectional Absolute Deviation and $R_{m,t}$ is the market return on day *t* or at minute m. Estimations are run for three frequencies (five-minute interval, ten-minute interval, and one-day interval). T-statistics are reported in brackets. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 4.3 Basic herding specification

	Five-minute frequency	Ten-minute frequency	One-day frequency
$(1 - D_t/m)R_{mkt,t/m}$	0.6687***	0.5962***	0.1694**
$D_{t/m}R_{mkt,t/m}$	-0.6997***	-0.6321***	-0.2746***
$(1 - D_{t/m})R_{mkt,t/m}^2$	-0.0148	0.0042	0.0924**
$D_{t/m}R_{mkt,t/m}^2$	-0.0482***	-0.0334***	-0.0315*
$D_{low,t/m}$	-0.0135***	-0.0113***	0.0025
$D_{up,t/m}$	-0.0070***	-0.001	0.0063

The table reports the results of the basic herding specification from the non-linear equation $CSAD_{t/m} = \alpha + \beta_1(1 - D_{t/m})R_{mkt,t/m} + \beta_2 D_{t/m}R_{mkt,t/m} + \beta_3(1 - D_{t/m})R_{mkt,t/m}^2 + \beta_4 D_{t/m}R_{mkt,t/m}^2 + \beta_5 D_{low,t/m} + \beta_6 D_{up,t/m} + \varepsilon_{t/m}$, where CSAD is the cross-sectional absolute deviation on day t or at minute m, D_t is a dummy variable that takes the value of one when the market return $R_{mkt,t}$ is negative and the value of zero otherwise, $D_{low,t}$ is a dummy variable that takes the value of one when the market return is located in the lower 5% tail of its distribution, and $D_{up,t}$ is a dummy that takes the value of one when the market return is located in the upper 5% tail of its distribution. Estimations are run for three frequencies (five-minute interval, ten-minute interval, and one-day interval) for sample period from January 2015 to December 2017. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 4.4 Herding under extreme values of HFT proxies

	Quote_update		Hft_volume	
	Low 5%	High 5%	Low 5%	High 5%
$(1-D_{t/m})R_{mkt,t/m}$	0.2595	0.3627	-0.4538	0.4226
	(0.61)	(0.95)	(-1.44)	(1.47)
$D_{t/m}R_{mkt,t/m}$	-0.9576	-0.098	0.8202	-0.0457
	(-1.36)	(-0.48)	(2.39)**	(-0.19)
$(1-D_{t/m})R^2_{mkt,t/m}$	0.1084	-0.1386	1.0453	-0.0850
	(0.11)	(-0.57)	(1.91)*	(-0.51)
$D_{t/m}R^2_{mkt,t/m}$	-3.2595	-0.0008	1.7219	0.0135
	(-1.20)	(-0.01)	(3.28)***	(0.18)
Adj R^2	0.15	0.09	0.39	0.26

This table presents the results of herding specification for various subsamples given by non-linear regression: $CSAD_{t/m} = \alpha + \beta_1(1-D_{t/m})R_{mkt,t/m} + \beta_2 D_{t/m}R_{mkt,t/m} + \beta_3(1-D_{t/m})R^2_{mkt,t/m} + \beta_4 D_{t/m}R^2_{mkt,t/m} + \varepsilon_{t/m}$, where CSAD is the cross-sectional absolute deviation on day *t* or at minute m, $D_{t/m}$ is a dummy variable that takes the value of one when the market return $R_{mkt,t/m}$ is negative and the value of zero otherwise. The samples include lower 5% tail and upper 5% tail of two HFT proxies: quote update and intensity from January 2015 to December 2017. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 4.5 Intentional herding and spurious herding

All stocks						
Dependent variable	Total CSAD ($CSAD_t$)		Non-fundamental driven CSAD ($CSAD_{NONFUND,t}$)		Fundamental driven CSAD ($CSAD_{FUND,t}$)	
HFT Proxies	β_1	β_2	β_1	β_2	β_1	β_2
Quote_update						
Low 5%	0.0018	0.0016	-0.00001	0.0038	0.0018	-0.0022
	(0.46)	(0.18)	(-0.01)	(0.52)	(0.74)	(-0.40)
Quartile 1	0.0003	0.0026	-0.0001	0.0033*	0.0004*	-0.0007
	(0.33)	(1.38)	(-0.09)	(1.79)	(1.81)	(-1.57)
Quartile 2	0.0012	-0.0003	0.0011	-0.0006	0.00003	0.0002
	(1.15)	(-0.26)	(1.15)	(-0.46)	(0.06)	(1.29)
Quartile 3	0.0007	0.0006	0.0003	0.0009	0.0004**	-0.0003*
	(0.85)	(0.82)	(0.38)	(1.29)	(2.04)	(-1.83)
Quartile 4	0.0006	0.0003	0.0005	0.0003	0.00002	-0.00008
	(0.71)	(0.59)	(0.75)	(0.82)	(0.07)	(-0.44)
High 5%	0.0014	-0.0001	0.0006	-0.00003	0.0008	-0.0001
	(0.85)	(-0.33)	(0.42)	(-0.06)	(0.92)	(-0.53)
Hft_volume						
Low 5%	-0.0059**	0.0134***	-0.0050*	0.0111**	-0.0009	0.0023
	(-2.12)	(3.17)	(-1.88)	(2.72)	(-0.62)	(1.08)
Quartile 1	0.0021***	-0.0002	0.0024***	-0.0007	-0.0003	0.0005*
	(2.69)	(-0.28)	(3.27)	(-0.94)	(-1.32)	(1.91)
Quartile 2	0.0018**	0.0006	0.0014*	0.0007	0.0003	-0.00003
	(2.25)	(0.87)	(1.80)	(0.93)	(1.49)	(-0.15)
Quartile 3	0.0021**	-0.0001	0.0013	0.0006	0.0008**	-0.0007***
	(2.29)	(-0.15)	(1.53)	(0.88)	(2.42)	(-2.69)
Quartile 4	0.0022***	-0.0001	0.0021***	-0.0002	0.00007	0.0001
	(3.35)	(-0.32)	(3.45)	(-0.76)	(0.25)	(0.86)
High 5%	0.0037**	-0.0007	0.0040**	-0.0008	-0.0002	0.00006
	(2.19)	(-1.33)	(2.51)	(-1.53)	(-0.59)	(0.40)

The table reports results from the non-linear regressions: $CSAD_{FUND,t} = \beta_0 + \beta_1 |R_{mkt,t}| + \beta_2 R^2_{mkt,t} + \varepsilon_t$, and $CSAD_{NONFUND,t} = \beta_0 + \beta_1 |R_{mkt,t}| + \beta_2 R^2_{mkt,t} + \varepsilon_t$. $CSAD_{NONFUND,t} = \epsilon_t$ is estimated from the OLS regression: $CSAD_t = \beta_0 + \beta_1 (R_{mkt,t} - R_f) + \beta_2 HML_t + \beta_3 SMB_t + \beta_4 MOM_t + \epsilon_t$; $CSAD_{FUND,t} = CSAD_t - CSAD_{NONFUND,t}$. The samples use two HFT proxies to split into four equal-sized quantiles (quantile 1 is the smallest; quantile 5 is the largest), as well as the lower 5% and upper 5%. Estimations are run for one-day frequency for the January 2015 to December 2017 sample period. T-statistics are reported in brackets. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Chapter 5 Colocation and Stock Herding

5.1 Introduction

Academic researchers have been paying increasing attention to examine investors' behaviours in equity markets on how they trade (e.g. Christie and Huang, 1995; Chang et al., 2000; Chiang and Zheng, 2010; Cui, Gebka and Kallinterakis, 2019; Gavriilidis et al., 2021). When investors discard or suppress their beliefs and mimic others' activities on trade regardless intentional or unintentional (spurious), such behaviour referred to as herding (Bikhchandani and Sharma, 2000). Previous literature mainly focuses on human trading which caused by human behaviour's bias, to examine the presence of herding in equity markets (e.g. Christie and Huang, 1995; Chang et al., 2000; Chiang and Zheng, 2010; Galariotis et al., 2015). However, the impact of non-human trading as algorithm trading and HFT on equity markets have not received enough attention on herding's literature.[21] In this empirical chapter, we use daily date of 10 exchanges around the world to test whether emergence of HFT can induce herding.

Spyrou (2013) reveals different topics of herding have been examined in equity markets. The findings of herding vary in different equity markets, as well as differ on unconditional herding (Christie and Huang, 1995; Chang et al., 2000; Gleason et al., 2004; Chiang and Zheng, 2010) and conditional herding (Galariotis et al., 2015; Bernales et al., 2020; Voukelatos and Verousis, 2018). Chapter 4 uses HFT as a condition and propose that herding could be no human related. Considering HFT is algorithm basis which follows computer routes and algorithm designs and apply similar trading strategies, they evidence HFT activity is not human related and therefore induce spurious herding instead of intentional herding in the US equity market.

[21] Andrikopoulos et al. (2017) use high frequency data to estimate herding. Chapter 4 also applying high frequency date to test for herding towards the market consensus for the US stock market and evidence HFT can induce herding.

Given Chapter 4 evidence HFT activity can induce herding in equity market, we focus on a more targeted research question in this empirical chapter. Instead of examining HFT activity, we investigate the emergence of HFT as a signal starting to affect exchanges around the world. There is no established HFT start date for exchanges, and therefore, we use two proxies (i.e. HFT effective date and colocation start date) to represent the emergence of HFT and to explore whether HFT emergence can explain herding on exchanges. Before exchanges formally offer colocation services to allow HFT to locate the servers in exchanges, HFT already started to move the servers as close as possible to exchanges, in order to increase trading speed and react promptly on information. Therefore, we use HFT effective date from Aitken et al. (2015) as our first proxy of emergence of HFT. The second proxy for emergence of HFT is colocation start date when exchanges formally offer colocation services (Boehmer et al., 2020).

Considering HFT is motivated by the algorithm strategies which following similar computer routes, HFT activities has advantage of high speed and react to the same information by using the similar methods to calculate the decision. Therefore, HFT activities unintentionally induce herding (result from Chapter 4). HFT compete to gain higher speed by locating the servers as close as possible to exchanges in order to achieve lower latency and be able to response to the trade before information merged into prices. Therefore, the introduction of colocation services from exchanges, which allows traders to locate their servers geographically closer to exchanges. Due to the competition among HFT, some might already rent a place closer to exchanges before the colocation services are available (Aitken et al., 2015). So, we consider both situations and allow both HFT effective date and colocation start date as a signal of HFT emergence. The fact that either way will increase trading speed of HFT and will increase the frequency to run the similar algorithm trade, therefore, we propose the emergence of HFT will induce herding on exchanges.

In this chapter, we investigate this issue by examining the role of emergence of HFT in the induction of herding, using daily data includes three years of pre-HFT effective date and three years of post-colocation date for constituents in indexes from ten exchanges in nine countries. We first investigate whether herding is significant to the effect of HFT emergence (proxy by colocation and HFT effective date). Secondly, we explore whether colocation or HFT effective date has stronger effect to induce herding under different market conditions (i.e. extreme, negative, and positive market return). Thirdly, due to many exchanges have HFT emergence around subprime crisis period between Jan 2008 to Apr 2011, we assess the effect of subprime crisis on herding. [22]

We contribute to the literature by connecting the gap between the impact of HFT emergence and herding on exchanges. Further to Chapter 4, we provide more evidence regarding HFT induces herding on exchanges from the perspective of non-human trading. Previous literature mainly considers to what extent non-HFT (i.e. human trading) will result in herding presence as investors' behaviour (e.g. Christie and Huang, 1995; Chang et al., 2000; Chiang and Zheng, 2010; Galariotis et al., 2015). In contrast, our empirical chapter attempt to understand if HFT effectively affect exchanges and/or through the introduction of colocation services from exchanges can induce herding on exchanges. Our empirical results indicate: 1) herding evidence absence without consider the effect of emergence of HFT; 2) calendar dummies (i.e., January and December) might play a part to trigger herding but these results are not consistent; 3) colocation tends to induce herding when the market return is negative; 4) when market return is relatively high, HFT effective date is the main reason to induce herding; 5) although both HFT effective date and colocation start date can induce herding, introduction of colocation is more appropriate to proxy HFT start date and has stronger effect to induce

[22] We follow the subprime crisis period as in Galariotis et al. (2015).

herding. In contrast, the effect of HFT effective date is delayed before exchanges formally introduce colocation services.

The remaining of this empirical chapter is structured as follows. In section 5.2, we review the theoretical and empirical literature for colocation services, define HFT effective date, and identify our hypotheses. We construct methodology in section 5.3 and describe data in section 5.4. In section 5.5, we discuss our empirical results and provide implications of our findings and in section 5.6 is our conclusion of this empirical chapter.

5.2 Literature Review of HFT Emergence[23]

Some research focus on one exchange within a specific short time, depending on the data that could be used to identify HFT trading activities. In order to understand the effects of HFT on stock exchanges, it is necessary to know the HFT start date in different exchanges. However, there is no established HFT start date corresponding to various exchanges around the world, which results in difficulties for empirical research of HFT. Therefore, previous literature use colocation service start date in exchanges as the proxy of HFT emergence (Aitken et al., 2017; Boehmer et al., 2020), while Aitken et al. (2015, 2017) propose HFT effective date as another proxy of HFT emergence.

5.2.1 What is Colocation Services?

Colocation refers to exchanges offers opportunity to high frequency traders (HFTs) and firms to locate their servers in the same building as the stock exchange servers. Thus, providing a higher speed to the flow of time-sensitive information (Brogaard, Hagströmer and Nordén, 2015). In order to increase the trading speed, exchanges around the world were creating or will create a huge data centre which allows member, non-member, and traders to place computers with algorithm trading next to the exchanges to match 'buy' and 'sell' orders. [24] HFT firms also known to physically relocate their servers as close as possible to exchanges in order to achieve microsecond speed advantage (Aitken et al., 2017). HFTs as trader is the main "user" to receive colocation services directly from one or more exchanges (NYSE exchanges, 2020).

Some traders in the exchanges are looking for the best price and lowest latency to act faster than the competitors. When the computer with algorithm trading is far away from the exchanges, it will face microseconds latency when it tries to trade a price on the computer

[23] See comprehensive literature review in Chapter 2.
[24] Based on the description from NYSE exchange, 'member' indicates a trading permit holder or an owner with trading license who approved to trade on one or more of the NYSE markets. In contrast, 'non-member' represents any entity that without the characteristics of 'member'.

screen. Due to the automatic algorithm trading processes, few traders can afford the latency while prices change so quickly. If a short-term informed trader has the latency advantage, this trader can trade rapidly on orders which do not reflect the latest news information. Instead, the slow market maker cannot avoid adverse selection costs. Requirements of greater capacity and higher efficiency of trading time are becoming more and more important. Therefore, colocation service is a popular solution for those who can afford it. It allows traders' servers to locate as close as possible to matching engines of exchanges (Aitken et al., 2017). By linking the high-speed server to the exchanges, each mile closer will reduces around eight microseconds of latency (LSE, 2009). Traders are willing to pay any expenditure to diminish those microseconds, as they can make so much money through lowest latency.

5.2.2 Empirical Results of Colocation Services

Previous studies reveal some effects of colocation service in exchanges. Boehmer et al. (2020) use colocation service announcements as an indicator of algorithm trading in 22 exchanges to examine international differences and liquidity of algorithm trading. They indicate market liquidity is increased while short-term volatility is also increased during the 12 months after introducing colocation services, because colocation results in more intensive algorithm trading. Similarly, Frino, Mollica and Webb (2014) find liquidity has improved by introducing colocation service to futures traders in Australia exchanges. After introducing colocation services, the bid and ask spread declines while market depth increases, which provide evidence for the increase in HFT activities. Brogaard et al. (2015) indicate optional colocation upgrade service from NASDAQ OMX Stockholm exchange allows HFT to reduce the cost of liquidity provision, which leads to higher trading profit. Also, they conclude speed could increase market liquidity, even for non-colocated trading entities. However, they argue that no observation regarding changes of volatility after introducing colocation services, which is in line with the results from Chaboud et al. (2014).

Aitken et al. (2015) manually collected colocation start date from 22 exchanges as a proxy of HFT start date and show the presence of HFT has significantly reduced the frequency and severity of end-of-day price dislocation, and evidence HFT acts as a valuable role to facilitate price discovery. Based on their research, Aitken et al. (2017) added two more exchanges and use colocation started date as the proxy of algorithm trading and HFT and find mixed results for average trade size influence. They conclude HFT exists on most of the exchanges by at least 8 months before colocation services introduced.

In order to attract more algorithm trading activities, exchanges amend the market structure by introducing colocation services which allowed algorithm traders collocate their servers in the market data centre. Therefore, HFT which benefits from colocation service could quickly adjust their quote over the changes of market condition. From the perspective of exchanges, providing colocation services would also assure the competitive environment between HFT market makers. Therefore, the introduction of colocation services would increase competition among HFT market makers and increase intensity of HFT activities (Hendershott et al., 2011).

Baron et al. (2019) confirm that approximately half of HFT firms use colocation to reduce latency and react to market events. They examine two colocation upgrade events from NASDAQ OMX Stockholm exchanges and evidence HFT firms with faster speed have better trading performance. If all HFT subscribes colocation services, the introduction could reflect the speed upgrade of exchanges as colocation allows HFT enter into exchanges by reducing the path length to the matching machine, and therefore provide evidence of market liquidity improvement (Menkveld and Zoican, 2017).

5.2.3 Using Trade Size to Define HFT Effective Date

Aitken et al. (2017) test whether the emergence of HFT results in the introduction of colocation service and conclude colocation services of most exchanges caused by HFT. Therefore, we use HFT effective date from Aitken et al. (2015) to identify the start date of HFT.[25]

HFT effective date usually appears few years before colocation date. This probably due to HFT established themselves at locations proximate to exchanges earlier before colocation service offered (Aitken et al., 2017). Aitken et al. (2015) denote the trade size ($TSize_{i,t}$) as monthly trading volume divided by the monthly number of trades for the exchange. If there are four continuously months trading size declined or the biggest single drop from the previous month on an exchange, the first month or the biggest drop month as HFT starts to affect the market. They exclude the significant declines to be defined as HFT effective date while during the financial crisis period between 2007 and 2008.

Hypothesis I: HFT effective date induces herding.

5.2.4 Why HFT Emergence might Induce Herding?

Colocation introduction marks an event that is fairly homogenous across exchanges. Because this event specifically offers infrastructure to HFT activity and indicates a commitment of exchanges to accommodate such trading activity. We focus on colocation services started date from 10 exchanges to examine the induction of herding by the following reasons. First, HFT can generate faster speed to react to the macroeconomic announcements and reduce latency to adjust the quotes through colocation services. Given HFT is applying algorithmic trading strategy through computers, exchanges which offer colocation services would provide a better opportunity to HFT in order to program faster and run algorithm faster. As in Chapter 4 indicates, the spurious herding caused by HFT activities is possibly due to the computer routes

[25] There are eleven out of twenty-one exchanges already have HFT effective date available from Aitken et al. (2015). We will examine ten exchanges in this empirical chapter, especially for those exchanges without HFT effective date available from Aitken et al. (2015).

and algorithm designs are similar for HFT activities, and therefore faster speed through colocation services would increase the chances of HFT to run the same algorithm in an overlapping microsecond.

Second, exchanges that introduce colocation service would facilitate the improvement of HFT activities intensity (Boehmer et al., 2020; Frino et al., 2014), which means to attract more intensive algorithm trading activities (Hendershott et al., 2011). In Chapter 4, we confirm that high HFT intensity induces herding in the US equity market.[26]

Hypothesis II: Introduction of colocation service induces herding.

[26] Relevant literature review of herding is in Chapter 2.

5.3 Data

We collect our main data from the Thomson Reuters Eikon database, which contains daily closing prices of main index constituents for ten exchanges in nine countries, including Australia, Canada, Germany, Japan, India (Bombay and NSE), Sweden, Switzerland, United Kingdom, United State. The index constituents are updated based on years for each exchange. Regarding to the colocation date and HFT effective date for each exchange, we follow Boehmer et al. (2020) and Aitken et al. (2015), respectively. Considering the HFT effective date always earlier than the colocation date, our original data covers 3 years before the HFT effective date to 3 years after the colocation date for each exchange. The date varies according to the exchanges, and therefore the dataset that we used is an unbalanced panel data. The list of colocation start date and HFT effective date of each country is included in Table 5.1.

[Table 5.1 around here]

We calculate market return by using the daily closing prices of constituents corresponding to the main index of exchanges. We further create several dummy variables as follow: 1) January and December dummy variables which take value of one in the corresponding period, otherwise zero;[27] 2) crisis dummy variable equals one between Jan 2008 to Apr 2011, otherwise zero; 3) in order to capture the over-time effects from colocation date and HFT effective date, we generate two dummy variables which equal zero before colocation date and HFT effective date, then equal accumulated days after the start date.[28]

[27] We include these two calendar dummies as in Aitken, Cumming and Zhan (2015).
[28] Following the methodology of Boehmer, Fong and Wu (2012).

5.4 Methodology

5.4.1 Detecting Herding from Exchanges

We again follow Chang et al. (2000)'s method as Eq. (3.7) to estimate herding towards to the market consensus in the equity market based on the cross-sectional dispersion of stock returns around the market return. [29]

As we discuss in the literature review, the crucial issue of this study is whether HFT tends to induce herding after the exchanges introduces a colocation service. We introduce colocation start date as a dummy variable. We augment Eq. (3.7) as follows:

$$CSAD_{i,t} = \beta_0 + \beta_1 \left| R_{mkt,i,t} \right| + \beta_2 R_{mkt,i,t}^2 + \beta_3 Col_{i,t} R_{mkt,i,t}^2 + \varepsilon_t \tag{5.1}$$

where $Col_{i,t}$ denotes an incremental colocation dummy variable in exchange i on day t which equals zero before colocation date, then equals accumulated days after the colocation date.[30]

In order to control for other economic effects from market, we follow Aitken et al. (2017) to include the following dummy variables:

$$CSAD_{i,t} = \beta_0 + \beta_1 \left| R_{mkt,i,t} \right| + \beta_2 R_{mkt,i,t}^2 + \beta_3 Col_{i,t} R_{mkt,i,t}^2 + \beta_4 Jan_{i,t} R_{mkt,i,t}^2 + \beta_5 Dec_{i,t} R_{mkt,i,t}^2 + \beta_6 Crisis_{i,t} R_{mkt,i,t}^2 + \varepsilon_t \tag{5.2}$$

where $Jan_{i,t}$ is a dummy variable equals to 1 if the month is January in exchange i on day t and zero otherwise, $Dec_{i,t}$ is a dummy variable equals to 1 if the month is December in exchange i on day t and zero otherwise, $Crisis_{i,t}$ is a dummy variable equals to 1 if the year is 2008 in exchange i and zero otherwise. In order to test hypothesis one, we expect the statistically

[29] Review the detailed methodology in Chapter 3.
[30] We followed Boehmer et al. (2020) to create colocation dummy variable.

significant negative coefficient of colocation dummy variable to show the introduction of colocation service induces herding. We then use HFT effective date to replace colocation start date as a proxy of HFT start date, where $HFTstart_{i,t}$ is a dummy variable which equals zero before HFT effective date, then equals accumulated days after the start date.

We include HFT effective date to augment Eq. (5.2) as follows:

$$CSAD_{i,t} = \beta_0 + \beta_1 \left| R_{mkt,i,t} \right| + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t} R^2_{mkt,i,t} + \beta_4 HFTstart_{i,t} R^2_{mkt,t} + \beta_5 Jan_{i,t} R^2_{mkt,i,t} + \beta_6 Dec_{i,t} R^2_{mkt,i,t} + \beta_7 Crisis_{i,t} R^2_{mkt,i,t} + \varepsilon_t \tag{5.3}$$

5.5 Empirical Results

We list 10 exchanges in 9 countries in Table 5.1 which also presents colocation date and HFT effective date of each exchange. The Germany and America exchanges' colocation date started before the subprime crisis period, while that of other countries started after 2008. In Table 5.2, we present the descriptive statistics for the full sample dataset. The summary statistics show different market return from day to day in different exchanges.

[Table 5.2 around here]

Table 5.3 presents results for Eq. (3.7) for the full sample and the sub-samples. For each sample period, we show the estimated coefficients and t-statistics for all days on the first line. The second line presents results for only those days with market returns greater than zero (i.e. 'up' days), while the third line shows the results on those days with negative market returns (i.e. 'down' days). The first and third columns present the estimated coefficients for absolute market return and market return square, respectively. The second and fourth columns show the respective t-statistics. As the results indicate, the coefficients on the non-linear term are all positively significant which evidence herding is not for either period.

[Table 5.3 around here]

We present the results for extreme market returns in Table 5.4 by applying Eq. (5.3). The values of β_1 are positive and statistically significant for all extreme market returns, implying no herding but return dispersion deviates from rational pricing model during period of large price movements. The first column shows the results for lowest one percent of market return, which indicates herd behaviour is present as the estimated coefficient for non-linear term is negatively significant. But it is not due to the introduction of colocation. The negatively significant coefficient in second column evidence the introduction of colocation is the reason to trigger herding for lower five percent market returns. The results of market return on the top five percent and one percent are consistently significant and negative on non-linear term of

HFT effective date, which indicate the herd behaviour under extremely high market return is due to the beginning of HFT.

[Table 5.4 around here]

In Table 5.5, β_1 shows similar results as in Table 5.4 except the result for quintile 3. The results in Table 5.5 indicate that the emergence of HFT induces herding. We split four quintiles across the 25, 50, and 75 percentiles of the distribution of market return. The first column presents that the introduction of colocation and December effect induce herding when the market return is low. Statistically significant and negative estimated coefficients are the proof of this. The second and fourth column present the opposite results where colocation no longer play a role on herd behaviour, instead, the estimated coefficients of HFT effective date are negative and statistically significant on 1% level. This suggests the medium-low and high market returns are firstly affected by HFT effective date which is earlier than the introduction day of colocation in the exchanges. The results in the column three are mixed. It indicates herding behaviour is present under the medium-high market returns, which induces by the HFT effective date and enhanced by the influence of colocation introduction.

[Table 5.5 around here]

Table 5.6 illustrates the regression results during subprime crisis period, where it starts on Jan 2008 and ends on Apr 2011. [31] Model 1 shows the relationship directly between CSAD and market return. Model 2 and Model 3 test the results with HFT effective date dummy and colocation dummy, respectively. Model 4 and Model 5 are the regressions testing the positive market return and negative market return with additional calendar dummies, respectively. Model 6 shows the joint results for the full sample during subprime crisis period.

Herd behaviour absence during subprime crisis period if we only consider the market return in Model 1. However, we find HFT effective date plays an important role to induce

[31] We follow Galariotis et al., (2015) to define the crisis period.

herding during subprime crisis period, where Model 2, 4, and 5 present negative and statistically significant coefficient. We find no herding evidence through introduction of colocation itself, but with the influence of HFT effective date. Specifically, colocation introduction induces herding when the market return is poor, which in line with our previous results. Colocation introduction date and HFT effective date are both inducing herd behaviour when the market return is positive during subprime crisis period. Coefficients of January dummy are positive and significant, indicating that January effect will result in the higher return dispersion but will not induce herding.

[Table 5.6 around here]

Table 5.7 presents the results of whether HFT effective date induces herd behaviour even before colocation introduced. We exclude colocation dummy as we test only the period before the exchanges introduce colocation, which include the period between HFT effective date and colocation start date. We find herding evidence on positive market return, negative market return, and the joint variables' results. However, these results show that herd behaviour is not induced by HFT effective date, because none of the estimated coefficients of HFT effective date are negative and statistically significant. The presence of herding is due to other reasons. For example, December effect under the negative market return can explain the presence of herding.

[Table 5.7 around here]

In Table 5.8, we test whether both HFT effective date and colocation start date can induce herding. Our sample range takes post HFT effective date, which also includes post colocation start date as introduction of colocation always starts later than the HFT effective date. Model 1 tests basic herding with HFT effective date dummy. Models 2 replaces HFT effective date with colocation dummy. Model 3 and Model 4 estimate positive market return and negative market return with additional calendar dummies, respectively. Model 5 tests all joint variables.

The results from Table 5.8 show the mixed results of HFT effective date and colocation date. Both dummies have independent effect to induce herd behaviour, where estimated coefficients of HFT effective date and colocation date are negative and statistically significant in Model 1 and Model 2. HFT effective date tends to induce herding under the positive market return, while introduction of colocation induces herding when the market return is negative. The joint variables' results show colocation introduction is the main reason to induce herding, where the estimated coefficient is negative and statistically significant on 1% level. Meanwhile, subprime crisis is also a reason to enhance herd behaviour. Eight out of ten exchanges start to introduce colocation service after 2008. Considering HFT effective date can induce herding after we include the effect of colocation introduction, it might be the influence of HFT effective date is delayed or introduction of colocation is more likely to induce herding.

[Table 5.8 around here]

To test whether HFT effective date or colocation start date is more appropriate to proxy HFT start date, we perform 7 different models for the full sample and show the results in Table 5.10. Model 1 and Model 2 test whether HFT effective date and colocation start date can induce herd behaviour, respectively. Model 3 and Model 4 show the results after controlling subprime crisis and after controlling calendar dummies corresponding to Model 1 and Model 2. Model 5 and Model 6 show results for positive market return and negative market return, respectively. Model 7 tests all variables jointly.

The results in Table 5.9 consistently show estimated coefficients are negative and statistically significant for both HFT effective date and colocation start date, indicating both induce herd behaviour. Results are consistent by controlling additional calendar dummies and subprime crisis. This further indicates HFT effective date and colocation start date are the reasons to explain herd behaviour. The estimated coefficients of HFT effective date and colocation start date are statistically negatively significant for positive market return on 1% level and 5% level, respectively. It indicates both proxies of HFT start date are jointly induced herd behaviour. However, the results in Model 6 are in line with our previous results that herding only induces by introduction of colocation when market returns are negative. In our last estimation (Model 7), our results show the estimated coefficient of colocation start date is negative and statistically significant on 1% level. Although both HFT effective date and colocation start date show the evidence to induce herd behaviour, introduction of colocation is more appropriate to proxy HFT start date.

[Table 5.9 around here]

5.6 Conclusion

In this empirical chapter, we included 10 exchanges around the world in our sample and introduced HFT effective date and colocation start date as the proxies for HFT emergence to examine the presence of herding on these exchanges. Our results indicate both proxies have strong power to explain the herding.

We test both HFT effective date and colocation start date as proxies of HFT start date and the results are consistently indicating emergence of HFT can induce herding behaviour. As we mentioned before, emergence of HFT could affect the exchanges even before the colocation services offered by exchanges. The coefficients of HFT effective date are negatively statistically significant especially on positive market returns, medium-high market returns, highest 5% market returns, and extremely high 1% market returns. These indicate HFT effective date tends to induce herding under these circumstances where the market return is optimistic. Introduction of colocation has jointly effect with HFT effective date to induce herding when market returns is medium-high, positive market return during subprime crisis, and positive market return on full sample range. In addition, colocation start date consistently leads to herding in exchanges when market returns are negative, on lowest 5% market returns, and market returns under 25% across the full sample. This could be due to introduction of colocation services increase the intensity of HFT activities, therefore, HFT becomes more active regardless uptrend or downtrend market.

Subprime crisis does not trigger herd behaviour, instead, HFT effective date can explain herd behaviour during subprime crisis period. Eight out of ten exchanges have effective HFT activities before subprime crisis started. However, when we test the results before exchanges introduce colocation services, HFT effective date is not responsible for the presence of herding. This might be the reason of the delayed effect of HFT activities on exchanges before they formally introduce colocation services to allow HFT physically and geographically located the servers. Our further results on post HFT effective date and full sample indicate both HFT effective date and colocation start date can induce herding. But introduction of colocation services from exchanges is when emergence of HFT starts to induce herd behaviour, along with the delayed effect from HFT effective date.

Overall, our results support prior results from Chapter 4 that HFT induces herding on exchange. We complement these results by evidencing the presence of herding from more exchanges around the world and further conclude the emergence of HFT can induce herd behaviour.

5.7 Tables

Table 5.1 HFT effective date and colocation start date

Country	Exchange Name	HFT Effective Date	Colocation Start Date	Data Period
Australia	Australia stock exchange	Apr2006	Nov2008	01May2003-31Oct2011
Canada	Toronto stock exchange	May2005	Nov2008	01Jun2002-31Oct2011
Germany	XETRA Germany	Jan2003	Q42006	01Feb2000-30Sep2009
India	Bombay stock exchange	May2009	15Nov2010	01Jun2006-14Nov2013
India	NSE India	May2009	Aug2009	01Jun2006-31Jul2012
Japan	Tokyo stock exchange	May2005	May2009	01Jun2002-30Apr2012
United Kingdom	London stock exchange	Feb2006	Sep2008	01Mar2003-31Aug2011
United State	NASDAQ	Jan2003	Apr2005	01Feb2000-31Mar2008
Sweden	Stockholm stock exchange	Apr2005	25Jun2008	01May2002-24Jun2011
Switzerland	Swiss stock exchange	Jan2004	24Jun2008	01Feb2001-23Jun2011

This table includes the list of 10 exchanges and the corresponding country. HFT effective dates are from Aitken et al. (2015), and colocation start dates are from Boehmer et al. (2020).

Table 5.2 Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
$R_{mkt,t}$	21,780	0.0002	0.0186	-0.6369	0.7232
$\lvert R_{mkt,t} \rvert$	21,780	0.0116	0.0145	0	0.7232
$R^2_{mkt,t}$	21,780	0.0003	0.0047	0	0.523
CSAD	21,780	0.0152	0.0363	0	1.5645
colocation	22,068	138.4838	230.223	0	783
HFT effective date	22,068	543.0069	543.638	0	1949
January	22,068	0.082	0.2744	0	1
December	22,068	0.0805	0.2721	0	1
crisis	22,068	0.3383	0.4731	0	1

Notes: This table shows statistics for the full sample dataset of daily data. The data range is various on each exchange from three years before HFT effective date to three years after colocation start date. Colocation and HFT effective date refer to dummy variables which equal zero before colocation date and HFT effective date, then equals accumulated days after that. January and December are seasonal dummy variables. Crisis is a dummy variable which equals one during the crisis period and zero otherwise.

Table 5.3 Testing basic herding towards the market consensus

Sample	R_{mkt}	R^2_{mkt}
Full sample		
All days	0.6719***	1.6949***
Up days (R_{mkt}>0)	0.6514***	1.3709***
Down days (R_{mkt}<0)	0.6554***	2.2946***
Pre-HFT effective date		
All days	-0.0701**	12.5749***
Up days (R_{mkt}>0)	0.088	11.6365***
Down days (R_{mkt}<0)	-0.655***	19.9018***
HFT - Colocation		
All days	0.7648***	1.6306***
Up days (R_{mkt}>0)	0.3363***	4.614***
Down days (R_{mkt}<0)	0.963***	1.2937***
Post-Colocation		
All days	0.5324***	1.42***
Up days (R_{mkt}>0)	0.3513***	1.3616***
Down days (R_{mkt}<0)	-0.0879	10.0209***

We present unbalanced panel results in this table in four different period. Besides the full sample, we split the full sample period into pre-HFT effective date, HFT-Colocation ranges between HFT effective date and colocation start date, and post colocation start date. 'Up days' denotes to days with positive market return (R_{mkt}>0), while 'Down days' denotes to days with negative market return (R_{mkt} <0). We follow non-linear regression: $CSAD_{i,t} = \beta_0 + \beta_1|R_{mkt,i,t}| + \beta_2 R^2_{mkt,i,t} + \varepsilon_{i,t}$, where $CSAD_{i,t}$ is the Cross Sectional Absolute Deviation for exchange i on day t and $R_{mkt,\text{i},t}$ is the market return. A negative and statistically significant coefficient of R^2_{mkt} suggests herding toward the market consensus. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 5.4 Herding under extreme market return

Variable	Down 1%	Down 5%	Up 5%	Up 1%
$\lvert R_{mkt,t} \rvert$	11.2207***	5.9719***	4.4674***	10.3117***
	(1.4097)	(0.299)	(0.2993)	(0.6695)
$R^2_{mkt,t}$	-54.8971***	-43.6717***	-31.1646***	-53.1842***
	(4.8706)	(2.2575)	(2.843)	(4.4246)
Colocation	0.0087	-0.0108**	0.0206*	0.0295**
	(0.0045)	(0.0065)	(0.012)	(0.0111)
HFT effective date	-0.0098	0.0011**	-0.0208***	-0.0206***
	(0.0027)	(0.0032)	(0.0079)	(0.0067)
January	40.9909***	37.3485***	33.9087***	38.5789***
	(4.2811)	(1.9776)	(2.0481)	(2.8393)
December	-3.7034	-4.9828	0.6276	-0.6387
	(10.8937)	(1.5408)	(2.9953)	(2.4951)
Crisis	-4.2922	-1.1571	13.5699***	10.7093***
	(4.4275)	(1.6807)	(1.9975)	(2.7757)

In this table, we control other calendar dummies on the lowest 1%, lowest 5%, highest 5%, and highest 1%. The results are given by non-linear regression: $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t} R^2_{mkt,i,t} + \beta_4 HFTstart_{i,t} R^2_{mkt,i,t} + \beta_5 Jan_{i,t} R^2_{mkt,i,t} + \beta_6 Dec_{i,t} R^2_{mkt,i,t} + \beta_7 Crisis_{i,t} R^2_{mkt,i,t} + \varepsilon_{i,t}$, where $Col_{i,t}$ denotes an incremental colocation dummy variable in exchange *i* on day *t* which equals zero before colocation date, then equals accumulated days after the colocation date, $HFTstart_{i,t}$ is a dummy variable which equals zero before HFT effective date, then equals accumulated days after the start date, $Jan_{i,t}$ and $Dec_{i,t}$ are the dummy variables take value of one in January and December, respectively, otherwise zero; $Crisis_{i,t}$ is the subprime crisis dummy variable equals one between Jan 2008 and Apr 2011, otherwise equals zero. All models control for country and year fixed effects. Standard errors are reported in parentheses. A negative and statistically significant coefficient of non-linear term suggests herding toward the market consensus. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 5.5 Testing herding on different percentiles

Variable	Quintile 1	Quintile 2	Quintile 3	Quintile 4
$\lvert R_{mkt,t}\rvert$	2.0857***	1.2503***	0.2648	2.0446***
	(0.5698)	(0.6119)	(0.4001)	(0.6369)
$R_{mkt,t}^2$	-28.2144***	-180.3833***	1.6494	-21.0992***
	(1.067)	(39.2709)	(37.295)	(1.1514)
Colocation	-0.0301***	0.0796	-0.1762*	-0.0009
	(0.0089)	(0.0982)	(0.0401)	(0.0138)
HFT effective date	0.0108***	-0.1356***	-0.0423**	-0.0152***
	(0.0025)	(0.0446)	(0.019)	(0.0038)
January	27.7542***	3.0646	72.3371***	22.6195***
	(0.9827)	(19.2849)	(17.9704)	(0.9125)
December	-3.6358**	-38.1745	-45.7567***	2.5256
	(2.3622)	(13.4747)	(8.4736)	(2.3735)
Crisis	3.9706***	75.0962***	76.0704***	8.6475***
	(0.8233)	(17.9308)	(13.9808)	(0.9669)

This table controls other calendar dummies. Quintile 1 refers to market returns lower than 25% percentile. Quintile 2 refers to market returns between 25% and 50% percentile. Quintile 3 refers to market returns between 50% and 75% percentile. Quintile 4 refers to market returns greater than 75% percentile. The results are given by non-linear regression: $CSAD_{i,t} = \beta_0 + \beta_1\lvert R_{mkt,i,t}\rvert + \beta_2 R_{mkt,i,t}^2 + \beta_3 Col_{i,t}R_{mkt,i,t}^2 + \beta_4 HFTstart_{i,t}R_{mkt,t}^2 + \beta_5 Jan_{i,t}R_{mkt,i,t}^2 + \beta_6 Dec_{i,t}R_{mkt,i,t}^2 + \beta_7 Crisis_{i,t}R_{mkt,i,t}^2 + \varepsilon_{i,t}$, where $Col_{i,t}$ denotes an incremental colocation dummy variable in exchange i on day *t* which equals zero before colocation date, then equals accumulated days after the colocation date, $HFTstart_{i,t}$ is a dummy variable which equals zero before HFT effective date, then equals accumulated days after the start date, $Jan_{i,t}$ and $Dec_{i,t}$ are the dummy variables take value of one in January and December, respectively, otherwise zero; $Crisis_{i,t}$ is the subprime crisis dummy variable equals one between Jan 2008 and Apr 2011, otherwise equals zero. Standard errors are reported in parentheses. All models control for country and year fixed effects. A negative and statistically significant coefficient of non-linear term suggests herding toward the market consensus. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 5.6 Testing herding during subprime crisis

Variable	Crisis	Crisis_HFT	Crisis_COL	Positive MR	Negative MR	All jointly
$\lvert R_{mkt,t} \rvert$	0.5007***	1.1762***	1.1676***	1.317***	0.7363***	1.2807***
	(0.0337)	(0.0505)	(0.0508)	(0.4483)	(0.3597)	(0.2948)
$R^2_{mkt,t}$	1.3418***	-6.7778***	-9.1904***	-2.2678**	-6.9685***	-4.2176***
	(0.0982)	(0.7137)	(0.6211)	(0.9145)	(1.344)	(0.7743)
Colocation			0.0004	-0.0029**	-0.0395***	0.0201***
			(0.0011)	(0.0199)	(0.0308)	(0.0127)
HFT effective date		-0.0036***		-0.0087***	0.015***	-0.0145***
		(0.0005)		(0.0053)	(0.0144)	(0.0052)
January		12.1033***	9.5801***	18.3812***	20.9361***	14.0321***
		(0.6704)	(0.6423)	(0.9205)	(1.1921)	(4.7866)
December		-0.5197	-1.2433	0.8329	-1.5075	1.4191
		(1.4231)	(1.423)	(1.6368)	(1.8661)	(1.3338)

This table presents the panel results for subprime crisis period. The results of first model are given by non-linear regression: $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \varepsilon_{i,t}$. The second and the third model are following $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t} R^2_{mkt,i,t} + \varepsilon_{i,t}$. β_3 use HFT effective date dummy in the second model, while the third model replaces to colocation dummy. Results for last three models are given by: $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t} R^2_{mkt,i,t} + \beta_4 HFTstart_{i,t} R^2_{mkt,t} + \beta_5 Jan_{i,t} R^2_{mkt,i,t} + \beta_6 Dec_{i,t} R^2_{mkt,i,t} + \varepsilon_{i,t}$, where $Col_{i,t}$ denotes an incremental colocation dummy variable in exchange i on day t which equals zero before colocation date, then equals accumulated days after the colocation date, $HFTstart_{i,t}$ is a dummy variable which equals zero before HFT effective date, then equals accumulated days after the start date, $Jan_{i,t}$ and $Dec_{i,t}$ are the dummy variables take value of one in January and December, respectively, otherwise zero. Standard errors are reported in parentheses. All models control for country and year fixed effects. A negative and statistically significant coefficient of non-linear term suggests herding toward the market consensus. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 5.7 Testing herding for pre-colocation start date

Variable	HFT	Positive MR	Negative MR	All jointly
$\lvert R_{mkt,t} \rvert$	0.7027***	1.0496***	1.5018***	1.5061***
	(0.0298)	(0.2484)	(0.4206)	(0.4115)
$R^2_{mkt,t}$	2.3856***	-23.2615***	-22.5906***	-25.0607***
	(0.0951)	(1.1492)	(0.9874)	(0.741)
HFT effective date	0.0018**	0.0002	0.0178***	0.0074***
	(0.0007)	(0.0048)	(0.0112)	(0.0067)
January		32.7134***	23.1766***	26.2558***
		(1.0453)	(0.945)	(0.7082)
December		0.651	-8.0743**	-3.0707
		(4.9521)	(2.2954)	(2.0303)
Crisis		18.1121***	1.8734**	11.4809***
		(1.0134)	(0.8817)	(0.671)

This table shows the panel results before exchanges start to offer colocation services. The results of first model are given by non-linear regression: $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 HFTstart_{i,t} R^2_{mkt,i,t} + \varepsilon_{i,t}$. Results for last three models are given by: $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t} R^2_{mkt,i,t} + \beta_4 HFTstart_{i,t} R^2_{mkt,t} + \beta_5 Jan_{i,t} R^2_{mkt,i,t} + \beta_6 Dec_{i,t} R^2_{mkt,i,t} + \varepsilon_{i,t}$, where $Col_{i,t}$ denotes an incremental colocation dummy variable in exchange *i* on day *t* which equals zero before colocation date, then equals accumulated days after the colocation date, $HFTstart_{i,t}$ is a dummy variable which equals zero before HFT effective date, then equals accumulated days after the start date, $Jan_{i,t}$ and $Dec_{i,t}$ are the dummy variables take value of one in January and December, respectively, otherwise zero, $Crisis_{i,t}$ is the subprime crisis dummy variable equals one between Jan 2008 and Apr 2011, otherwise equals zero. Standard errors are reported in parentheses. All models control for country and year fixed effects. A negative and statistically significant coefficient of non-linear term suggests herding toward the market consensus. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 5.8 Testing herding for post HFT effective date

Variable	HFT	COL	Positive MR	Negative MR	All jointly
$\lvert R_{mkt,t} \rvert$	0.6198***	0.6156***	1.1673***	1.3029***	1.3025***
	(0.0241)	(0.0241)	(0.0467)	(0.0577)	(0.0367)
$R^2_{mkt,t}$	1.8493***	1.8276***	-14.2822***	-22.9924***	-10.9749***
	(0.0922)	(0.0903)	(1.8332)	(1.0074)	(0.5784)
Colocation		-0.0015***	0.0006	-0.0333***	-0.0074***
		(0.0002)	(0.0024)	(0.0017)	(0.0011)
HFT effective date	-0.0004***		-0.0122***	0.0129***	0.0026***
	(0.0007)		(0.0009)	(0.0006)	(0.0004)
January			16.5901***	23.7476***	11.8994***
			(0.6651)	(0.9472)	(0.5301)
December			2.2763*	-2.8943	-1.4564
			(1.2334)	(2.1549)	(1.1685)
Crisis			13.7877***	3.8397***	-1.554***
			(1.6783)	(0.8566)	(0.4725)

This table presents the panel results for post HFT effective date. The first and second model are following $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t} R^2_{mkt,i,t} + \varepsilon_{i,t}$. β_3 use HFT effective date dummy in the first model, while the second model replaces to colocation dummy. Results for last three models are given by: $CSAD_{i,t} = \beta_0 + \beta_1 \lvert R_{mkt,i,t} \rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t} R^2_{mkt,i,t} + \beta_4 HFTstart_{i,t} R^2_{mkt,t} + \beta_5 Jan_{i,t} R^2_{mkt,i,t} + \beta_6 Dec_{i,t} R^2_{mkt,i,t} + \beta_7 Crisis_{i,t} R^2_{mkt,i,t} + \varepsilon_{i,t}$, where $Col_{i,t}$ denotes an incremental colocation dummy variable in exchange i on day t which equals zero before colocation date, then equals accumulated days after the colocation date, $HFTstart_{i,t}$ is a dummy variable which equals zero before HFT effective date, then equals accumulated days after the start date, $Jan_{i,t}$ and $Dec_{i,t}$ are the dummy variables take value of one in January and December, respectively, otherwise zero, $Crisis_{i,t}$ is the subprime crisis dummy variable equals one between Jan 2008 and Apr 2011, otherwise equals zero. Standard errors are reported in parentheses. All models control for country and year fixed effects. A negative and statistically significant coefficient of non-linear term suggests herding toward the market consensus. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 5.9 Testing herding on the full sample

Variable	HFT	COL	HFT_Crisis	COL_Crisis	Positive MR	Negative MR	All jointly
$\lvert R_{mkt,t}\rvert$	0.65***	0.645***	1.393***	1.418***	1.045***	1.316***	1.406***
	(0.019)	(0.019)	(0.028)	(0.028)	(0.039)	(0.04)	(0.028)
$R^2_{mkt,t}$	2.51***	2.441***	-12.177***	-13.068***	-7.852***	-20.359***	-12.69***
	(0.085)	(0.084)	(0.485)	(0.478)	(0.741)	(0.752)	(0.49)
Colocation		-0.003***		-0.004***	-0.005**	-0.029***	-0.007***
		(0.0002)		(0.0007)	(0.0024)	(0.001)	(0.001)
HFT effective date	-0.001***		-0.0005*		-0.01***	0.011***	0.001***
	(0.0007)		(0.0002)		(0.0009)	(0.0005)	(0.0004)
January			13.645***	14.493***	16.619***	21.258***	14.103***
			(0.444)	(0.435)	(0.613)	(0.708)	(0.449)
December			-1.141	-1.284	1.345	-3.907**	-1.501
			(1.019)	(1.017)	(1.151)	(1.755)	(1.019)
Crisis			-0.477	0.418	6.549***	3.745***	-0.568
			(0.392)	(0.273)	(0.704)	(0.6157)	(0.392)

This table presents the panel results for the full sample. The first and second model are following $CSAD_{i,t} = \beta_0 + \beta_1\lvert R_{mkt,i,t}\rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t}R^2_{mkt,i,t} + \varepsilon_{i,t}$. β_3 use HFT effective date dummy in the first model, while the second model replaces to colocation dummy. The third and the fourth model include additional calendar dummy variables. Results for last three models are given by: $CSAD_{i,t} = \beta_0 + \beta_1\lvert R_{mkt,i,t}\rvert + \beta_2 R^2_{mkt,i,t} + \beta_3 Col_{i,t}R^2_{mkt,i,t} + \beta_4 HFTstart_{i,t}R^2_{mkt,t} + \beta_5 Jan_{i,t}R^2_{mkt,i,t} + \beta_6 Dec_{i,t}R^2_{mkt,i,t} + \beta_7 Crisis_{i,t}R^2_{mkt,i,t} + \varepsilon_{i,t}$, where $Col_{i,t}$ denotes an incremental colocation dummy variable in exchange *i* on day *t* which equals zero before colocation date, then equals accumulated days after the colocation date, $HFTstart_{i,t}$ is a dummy variable which equals zero before HFT effective date, then equals accumulated days after the start date, $Jan_{i,t}$ and $Dec_{i,t}$ are the dummy variables take value of one in January and December, respectively, otherwise zero, $Crisis_{i,t}$ is the subprime crisis dummy variable equals one between Jan 2008 and Apr 2011, otherwise equals zero. Standard errors are reported in parentheses. All models control for country and year fixed effects. A negative and statistically significant coefficient of non-linear term suggests herding toward the market consensus. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Chapter 6 Herding on Different Traders

6.1 Introduction

In order to ensure fairer, safer and more efficient markets and facilitate greater transparency for all traders, the implementation of Markets in Financial Instruments Directive (hereafter MiFID) II enhances market regulation. The protection of investors is strengthened through the introduction of new requirements on product governance and independent investment advice, the extension of previous rules to structured deposits, and the strengthening of requirements in several areas.[32] In general, MiFID II relates to the framework of trading venues or structures in which financial instruments are traded across European Union (EU) financial markets. This increased competition among EU financial markets, strengthening supervisory powers, and ensuring appropriate levels of investor protection. More specifically, it requires to identify algorithmic trading and non-algorithmic trading (i.e. from human traders).

In this chapter, we investigate herding towards the market consensus for the most traded stocks with HFT execution data on Athens Stock Exchange from January 2018 to December 2020. This chapter contributes to the literature in several ways. Firstly, we address a gap in the literature by testing herding for both low-frequency traders (i.e. human traders) and high-frequency traders (i.e. who use algorithms) with the dataset which contains trading execution information for all these traders. We identify three types of trader (human traders, algorithmic traders, and market makers) and test for herding to each trader accordingly. To the best of our knowledge, it is the first time that this issue is addressed in the literature. Previous studies tend to investigate herding on the market as a whole and ignore effects from different traders. In this chapter, the sample period of our dataset is started from January 1, 2018, which is three days before MiFID II implemented. This allows us to identify the specific trader for trading

[32] Including on the responsibility of management bodies, inducements, information and reporting to clients, cross-selling, remuneration of staff, and best execution.

execution. More specifically, traders employ algorithmic strategies are expected to frequently run the pre-programme algorithms when they trade intensively, and therefore result in algorithm-related herding (not herd intentionally). Thus, we include dummy variables to capture the high intensity trading effects from each trader. Our results evidence that traders using algorithms (algorithmic traders and market makers) tend to herd when the trading intensity is high, while herding is absence for human traders.

Secondly, we decompose trades to execution on buy-side and execution on sell-side of each trader, to further test whether traders tend to herd in general or have greater propensity on a certain side of trading execution. Due to the similar reason as mentioned above, we control the high trading intensity for both sides of trade. Our findings indicate that market makers tend to herd regardless of trading on buy-side or sell-side, while herding evidence only presence on algorithmic sell side trading. Since MiFID II became effective, it brings more intense competition among traders because it allows traders easier and wider to access information (Guo and Mota, 2021). Algorithmic traders are informed traders because of the speed advantage allows them to react fast before the information incorporate into the price. Unlike its traditional counterpart, algorithmic market makers also compete to act on market prices sooner. Therefore, the informed algorithmic traders might flee when they suspect other more informed traders are present.

Thirdly, we sort all stocks based on market capitalization on December 31st of each sample year to investigate whether investment style (i.e. trade on small stocks vs. trade on large stocks) plays a part of inducing herding behaviour for traders. Previous literature conclude that herding is more likely to presence on extreme capitalization stocks. Our findings in line with previous studies and indicate that algorithmic traders and market makers tend to herd when they execute trade on small stocks and large stocks. More specifically, both traders show herding evidence of trading execution on large stocks when the trading intensity is high. The results are consistent after decomposed trades into buy-side and sell-side.

Our findings have implications for regulators and traders, meanwhile adding new evidence to herding literature after MiFID II implemented. We report significant herding for traders who use algorithms. Algorithmic traders would expect competitors who are also well-informed or even more sophisticated entered the market by employing similar algorithmic strategies. This might reduce trading profits for traders but increase the market quality. On the other hand, herding evidence is missing for human traders in this chapter. A possible explanation is that the implementation of MiFID II increased supervisory powers for traders, and the advantage of algorithmic traders over human traders has weakened.

The rest of this chapter is organized as follows. In Section 6.2, we review the relevant literature for algorithmic trading and empirical results related to MiFID II, respectively. In Section 6.3, we present research questions and hypotheses. Section 6.4 presents the data and the testing methodologies. Section 6.5 reports empirical results, while Section 6.6 concludes the chapter.

6.2 Literature Review[33]

6.2.1 Empirical Results of Algorithmic Trading

Algorithmic trading (AT) generally defined as the use of computer algorithms to automatically complete trades, submit orders, and manage these orders. In this chapter, we define AT as all traders who use algorithm to trade (Hendershott et al., 2011). Before AT became a popular trading method, fund managers who want to purchase shares might hire a broker-dealer to look for the counterparty and expect to execute all amount in a block trade. Alternatively, they split the orders to execute the trade, in order to not push up the price too much on the same trading day. AT results in more electronic trading, which is more convenient and cheaper for AT to copy and follow the behaviour of floor traders and lead to the decline activities of floor traders (Hendershott and Moulton, 2011).

AT can pre-set the time, price, amount, and route of trading. Also, AT can dynamically monitor the market conditions of different securities and trading venues and reduce market impact by optimizing orders through randomly decomposing large orders into several small orders. Different types of traders might use different algorithms to trade. For example, Jovanovic and Menkveld (2011) indicate some hedge funds and broker-dealers applying algorithm to provide liquidity and competing with the designated market maker or other liquidity provider. Statistical arbitrage funds use computer to fast process order flows of securities and fast process large amount of information in price changes, and then high frequently execute trading following the patterns in the data (Jovanovic and Menkveld, 2011). For assets which available to trade on many venues, the liquidity demander needs to use algorithmic smart order routers to decide where to send the order (Foucault and Menkveld, 2008). Engle et al. (2008) analyse algorithm execution data from Morgan Stanley to examine how the changes of algorithm aggressiveness will affect trading cost. By including the

[33] See comprehensive literature review in Chapter 2.

execution cost of buy-side institutions, Domowitz and Yegerman (2005) compare the performance among different algorithm providers. Overall, institution investors rely on algorithm to gradually trade a large number of stocks over time (Hendershott et al., 2011).

Traders' cognition limits can be mitigated through AT. Regardless the trade is centralized or decentralized, traders need to monitor the source of information, such as bid and ask spreads, placed orders, the time series of trades, macroeconomic announcements, order's book position and so on. If these data are manually analysed, it will delay the trade and thus will reduce the realization of profit. Under this circumstance, AT can compensate traders' limited cognition and limited rationality, and therefore improve market liquidity (Biais and Woolley, 2011). They test the situation when the market is under aggregate liquidity shock, most of investors temporary reduce their willingness to hold the assets. Asset can hardly reallocate to traders who value it the most because of traders' cognition is limited. However, AT can mitigate this imperfection of the market. For example, algorithmic traders can execute buy orders with a relatively low price on the early stage after the shock, meanwhile, they place limit orders to sell when the price rebound later.

Previous studies compare the performance between algorithmic traders and human traders (i.e. without using algorithm to trade), and indicate some advantages for AT. Chaboud et al. (2014) first evidence algorithmic trading (i.e. computer) and non-algorithmic trading (i.e. human) have an impact on the information efficiency of foreign exchange prices. By extending Foucault (1999)'s study, Hoffman (2014) examines the competition between algorithmic traders and human traders and conclude algorithmic traders can avoid the adverse selection due to the ability of faster reaction on news. But the impact of price for human traders and the impact on trading volume are not necessarily due to the introduction of algorithm. Overall, Hoffman (2014) emphasise human traders' profits are declined, and it is hard for human traders to take trading advantage after AT widely used in the market. This is because algorithmic traders have the speed advantage relative to human traders, which allow them to react faster to public information and create positive influence on price informativeness. Biais, Foucault and Moinas (2015) and Martinez and Rosu (2012) indicate algorithmic traders are the informed traders who can utilize information to place their orders. In fact, the existence of algorithmic traders makes asset prices more informative. Moreover, algorithmic traders help price discovery. Once there are price inefficiencies, AT will quickly correct them through published quotation transactions. By contract, Menkveld and Yueshen (2012) assume information is symmetric and argue the speed advantages of algorithmic traders cannot increase price informativeness, instead, AT will increase the cost of adverse selection.

Some researchers use proxies to estimate AT, while other researchers utilize data that can identify as AT. For example, in one of the earliest studies, Hendershott et al. (2011) use the electronic messages of the New York Stock Exchange after the implementation of Autoquote to proxy AT. Similarly, instead of using Autoquote, Boehmer et al. (2020) employ the first available colocation date of exchanges around the world to proxy AT and to determine the impact of AT activities on liquidity, short-term volatility, and the information efficiency of stock prices. Without using proxies of AT, Hendershott and Riordan (2013) collect one month AT data on 30 DAX stocks which traded on Deutsche Börse. They conclude algorithmic traders improved market liquidity by providing liquidity when liquidity is scarce and consuming liquidity when liquidity is sufficient. In this chapter, we use data from Athens Stock Exchange which identified AT.

6.2.2 Algorithmic Trading under the Regulatory Environment

In order to create a level playing field in the market, European Union (EU) launched the MiFID on November 1, 2007. In fact, MiFID allows different national exchanges to compete and encourage new markets to enter. It enhances information availability for traders through improves trading transparency, investors protection, and competition (Investment services and regulated markets - Markets in financial instruments directive (MiFID), 2022). More specifically, trading transparency can reduce information asymmetric through increasing speed of information dissemination and decreasing costs of information acquisition. Especially for those less experienced traders, increased transparency can allow them to acquire valuable information from other sophisticated investors. Also, trading transparency allow traders to access competitors' transactions and increase incorporation of company-specific information, and therefore increase price informativeness. Level playing field created by MiFID can increase market participation and improve market liquidity. In order to reduce the uncertainty of investors, MiFID requires traders to execute the best incoming market orders. These rules aim to protect investors to promptly and sequentially execute orders, and thus encourage them to participate in the market. This will further improve market liquidity and price informativeness. Moreover, fierce competition provides trading opportunities on different venues, which reduces the cost of execution and increases market liquidity. Aghanya, Agarwal and Poshakwale (2020) examine the influence of MiFID to price informativeness and liquidity on 28 EU countries. They evidence MiFID has greater impact on price informativeness for countries with weaker quality of regulation and enforced regulation (i.e MiFID) increases market efficiency.

In order to make European financial markets more flexible, transparent and investor-friendly, and to better respond to financial crises, MiFID II became effective on January 3, 2018 (MiFID II, 2022).[34] Before MiFID II was implemented, brokerage fees which include transaction payment and other bundled services payment are not transparent. Instead of directly paying for bundled services, it compensates by transaction commissions. Under this situation, investors who face these costs cannot identify whether these are services cost or trading cost.[35] Also, it will encourage fund managers to charge additional service fees that clients might not

[34] MiFID II is applicable to 31 countries in the European Economic Area (EEA), which includes 28 EU countries plus Iceland, Liechtenstein and Norway.
[35] Bundled services payment refers to such as cost of investment research and cost of advisory services.

need (Biffany, 2018). MiFID II can mitigate these issues by separating payments for research and transactions. More specifically, information provided to clients must separately and transparently indicate all costs and fees, including any third-party payments. Fang, Hope, Huang and Moldovan (2020) evidence market liquidity declined after MiFID II became effective. Guo and Mota (2021) indicate the amount of research analysts reduced in large companies when unbundled costs of research and costs of transaction under MiFID II. Because the requirement to unbundle costs will bring more intense competition among analysts. This results in withdraw of inaccurate analysts and the remaining analysts conducted more accurate research.

Before implemented MiFID II, even though bundle services will create benefit confliction among buyers, intermediaries, and sellers, it covers high fixed costs of information and allows more information production. This will allow traders to easier and wider access to information (Guo and Mota, 2021). However, the requirement to unbundle costs will further enhance trading transparency and increase competition, which not only reduced benefit confliction among traders but also led to under-provision of certain information. Greater transparency might weaken the information advantage of sophisticated investors, reducing trading profits, and restrain them from actively trading in the market (Boulatov and George, 2013; Aghanya et al., 2020).

6.3 Hypotheses

There are two main differences between AT (i.e. computer trading) and non-AT (i.e. human trading). First, AT processes and reacts on information faster than human trading. Second, AT behaviour may be more relevant than non-AT behaviour. Because AT trades through computers which needs pre-programmed and reacts similar based on the given signal (Chaboud et al., 2014). Using 2008 and 2009 NASDAQ data, Brogaard et al. (2014) conclude high frequency traders facilitate price effectiveness by trading in the direction of permanent price changes, especially during macroeconomic news released time.[36] Hirschey (2021) employs the similar data and indicates aggressive purchases of HFT can anticipate the future aggressive purchases of non-HFT. Both research indicate high frequency traders are informed traders using algorithm to trade.

Almost every large broker-dealer provides its institutional clients with a set of algorithms to help them to execute orders. When AT pursues desired position, it expects to apply similar algorithm strategies when it captures common information after macroeconomic announcements (Chordia et al., 2018). As we mentioned above, the implementation of MiFID II makes information becomes more transparent, which results in the advantage of AI prioritizing obtain information less obvious. Choi and Skiba (2015) evidence institutional investors herd more under high level of information transparency, which indicate herding behaviour is driven by similar fundamental information. When AT reacts on information by running the similar pre-programme algorithms, it might unintentionally induce algorithm-related herding. Therefore, our first hypothesis is as follows:

Hypothesis I: Algorithmic traders induce herding behaviour.

AT affects market quality. Boehmer et al. (2020) indicate more intense AT activities result in more narrow spreads and improve liquidity. When spreads are narrow, Hendershott

[36] HFT is a subset of algorithmic trading (not all algorithmic traders trade with high frequency).

and Riordan (2013) conclude AT is less likely to submit new orders or cancel orders, instead, it is more likely to initiate trades. MiFID II will lead to increased competition, thus increased AT intensity and initiate more trades. Given AT using computer to apply similar algorithms, more intense activities implied more trades have been executed through computer routes. Therefore, algorithmic traders unknowingly behave consistent and collectively act like a large trader. Due to the fact of herding is where one mimic others trading activities, it is reasonable to expect non-AT traders (i.e. human traders) increased trading activities when herding presence, So, the second hypothesis is created as:

Hypothesis II: Intense algorithmic trading induces herding behaviour among non-AT trade.

6.4 Data and Methodology

6.4.1 Data

We collect data for the period after the implementation of MiFID II. The dataset includes executed trading data of 73 most-traded stocks from Athens Stock Exchange (ASE) in Greece from January 1, 2018 to December 31, 2020. We do not need to proxy AT in this chapter, as our dataset can identify three traders corresponding to algorithmic trades (algorithmic traders and market makers) and non-algorithmic trades (human traders).[37] The dataset contains information for each trade such as executed trading price and a flag to indicate which trader (i.e. human trader, algorithmic trader, and market maker) executed buy-side order and sell-side order.[38] The time stamp of the original dataset is to the millisecond. Considering the liquidity of Athens Exchange market is relatively low compared with other exchange markets (e.g. NASDAQ and LSE), certain stocks might not have executed trade in a short interval. Therefore, we aggregate the dataset on daily basis to create a fair comparison across stocks. We compute logrithm return for each stock *i* as follows: $R_{i,t} = (\ln(trade\_price_t) - \ln(trade\_price_{t-1}))$ on each daily interval *t*. Then we use the logrithm return to estimate market return for each stock *i* on day *t* as $R_{mkt,t} = \sum_{without\ stock\ i} \frac{R_t}{72}$.

Since our dataset can identify the trade is executed by either human traders, algorithmic traders, or market makers, we further test our hypotheses based on executed trading intensity from different traders. We accumulate daily number of trades from buy-side and sell-side for each trader, then compute total number of trades of each trader by adding corresponding buy-side trades and sell-side trades. For example, if the trade flags the algorithmic trader that

[37] According to MiFID II, sometimes the market maker registers the trade as algorithmic trading while sometimes not. No matter the executed trade from market maker has been registered or not, the information from dataset can identify the trade is executed by market maker.

[38] In our raw dataset, if a buy order is executed by an algorithmic trader, against an execution of a sell order from human trader, then our dataset will flag a buy-side trading by algorithmic trader and flag a sell-side trading by human trader. Therefore, we can clearly identify each executed trade from a specific trader.

executes a buy order against the execution of a sell order from the human trader, then algorithm trader will increase one trade from buy-side while human trader counts an additional trade from sell-side. We identify high trading intensity for each trader as the top 10% of its distribution across the full sample and generate three dummy variables for each trader. For example, three dummy variables of human trader are as follows: 1) $Intense_{up10,t}$ equals one if total number of trades of human trader on a day is located on the top 10 percent tail of its distribution over three years' sample period; 2) $Buy_{up10,t}$ takes value of one if number of trades executed from buy-side of human trader is very intense and located on the top 10 percent tail of execution on buy orders through the sample period; 3) we let $Sell_{up10,t}$ equals one if human trader frequently (i.e. on the top 10 percent over the sample period) execute sell orders. Then we use the same name of these three dummy variables and generate new values according to the specific trader (i.e., algorithmic traders and market makers).

In order to test whether there are differences in herding behaviour towards size effect, we rank all stocks in our dataset based on market capitalization. Data of market capitalization of all stocks for each sample year on 31st December are obtained from Bloomberg. More specifically, all stocks are ranked by market capitalization on each sample year. We assigned stocks with market capitalization on the lowest 20 percent of all stocks as small stocks, while large stocks refer to stocks with the highest 20 percent market capitalization over all stocks. Given that we are using the most traded 73 stock on ASE, one may argue that it is unnecessary to sort stocks based on market capitalization and analyse small stocks and large stocks separately. However, there is a large dispersion in market values of all stocks in our database. For example, as on 31st December 2019, the smallest market capitalization is €13.5 million, and the largest stock has a market capitalization of €10.9 billion.

6.4.2 Methodology

As the detailed methodology that we applied in Chapter 3, we first calculate the return for each stock i of all stocks N and then calculate the daily difference between the stock's return ($R_{i,t}$) and the market return ($R_{mkt,t}$), then we estimate the relation between the Cross-Sectional Absolute Deviation (CSAD) of stock return and market return. The cross-sectional dispersion of this measure is showed in Figure 6.1 from January 2018 to December 2020. In general, the evolution of the CSAD over time is relatively stable besides three notable exceptions where deviations from the market consensus increased significantly. However, as we discussed above, the level of dispersion does not reflect whether traders tend to herd or not. More specifically, low CSAD levels not necessarily suggest that traders have great propensity to follow market consensus, and high CSAD levels also not necessarily indicate that traders will price individual stocks independently.

[Figure 6.1 around here]

In order to get a basic idea of whether herding presence in our dataset, we plot CSAD against the equally weighted market return for full sample period in Figure 6.2. The relationship indicates that we are likely to find herding evidence, because larger absolute market returns are associated with lower levels of CSAD.

[Figure 6.2 around here]

We further use the method as in Chang et al. (2000) to estimate the cross-sectional absolute deviation (CSAD) of stock returns around the market return and to detect herding behaviour as Eq. (3.7).

We compute the basic herding specification of stock return from Eq. (3.7) on daily interval for all stocks, small stocks, and large stocks, where CSAD is regressed only against market return. As Chang et al. (2000) indicate, herding evidence presence when the coefficient of the non-linear term is negative and statistically significant.

As mentioned in the second hypothesis, an important question is whether traders tend to herd on days when they execute trades with high intensity. In order to test this hypothesis, we augment Eq. (3.7) as follows:

$$CSAD_t = \beta_0 + \beta_1\left|R_{mkt,t}\right| + \beta_2 R^2_{mkt,t} + \beta_3 Intense_{up10,t} + \beta_4 Intense_{up10,t} R^2_{mkt,t} + \varepsilon_t \qquad (6.1)$$

$Intense_{up10,t}$ is a dummy variable that takes the value of one when the intensity of total daily trades (i.e. regardless execution of buy orders or sell orders) is located on the upper 10 percent tail of its distribution corresponding to human traders, algorithmic traders, and market makers. For total number of trades on each day, we count trading execution no matter from buy-side or sell-side as one trade for corresponding traders. [39]

We first estimate herding for full sample which includes all stocks traded by three traders, [40] then we further test sub-samples when each trader trades on small stocks and large stocks. [41] If the coefficient of the first non-linear term (β_2) is negative and statistically significant for the trader, then it evidences this trader tends to herd. Regarding our second hypothesis in the case of herding due to high trading intensity, the coefficient on the second

[39] See Data section for more discussion.

[40] The estimation is applied to each trader separately. For example, when we test herding specification for algorithmic traders, the dummy variable takes value of one when total number of trades of algorithmic traders on a day is on top 10 percent of all daily number of trades through the sample period.

[41] We identify stocks with the lowest 20 percent market capitalization as small stocks, while large stocks have the highest 20 percent market capitalization over all stocks.

non-linear term (β_4) will be negative and statistically significant. The value of β_3 is expected to zero for the absence of herding and expected to be statistically significant if herding is present.

We next examine whether traders have greater propensity to herd when they execute buy order or execute sell order, especially when they trade intensively.

$$CSAD_t = \alpha + \beta_1|R_{mkt,t}| + \beta_2 R_{mkt,t}^2 + \beta_3 Buy_{up10,t} + \beta_4 Buy_{up10,t} R_{mkt,t}^2 + \varepsilon_t \quad (6.2)$$

$$CSAD_t = \alpha + \beta_1|R_{mkt,t}| + \beta_2 R_{mkt,t}^2 + \beta_3 Sell_{up10,t} + \beta_4 Sell_{up10,t} R_{mkt,t}^2 + \varepsilon_t \quad (6.3)$$

where $Buy_{up10,t}$ is a dummy variable that takes the value of one when trade execution intensity for orders from buy-side is on the top 10 percent of its distribution, instead, if trade execution is intensively from sell-side then $Sell_{up10,t}$ equals one.

Therefore, we split number of trades on a day into number of trades to buy and number of trades to sell. Similarly, we test all stocks before we examine sub-samples of small stocks and large stocks. We expect the coefficient of non-linear term (β_4) is negative and statistically significant to evidence high trading intensity (from buy-side or sell-side) can induce herding.

6.4.3 Descriptive Statistics

Table 6.1 reports information of descriptive statistics for our dataset from January 2018 to December 2020. In Panel A, we present some key statistical measures for Cross-Sectional Absolute Deviation and market return for full sample with all stocks, stocks with small capitalization, and stocks with large capitalization. It presents that the average CSAD for small stocks is higher than for large stocks, while small stocks' market return is lower than large stocks. Panel B outlines the same key statistical measures but for daily number of trades of three traders. Over the sample period, human traders have the highest average daily trading volume with approximately six times more than other traders. Algorithmic traders are relatively volatile. When they are actively executing trades, their maximum daily number of trades is one-third higher than market makers, otherwise their daily minimum trading volume is one-third less than that of market makers. Panel C presents the similar statistics by further classifying daily trading volume for each trader into buy-side execution volume and sell-side execution volume. Human traders and market makers are more interested in executing buy-side orders, while algorithmic traders prefer the execution of sell-side orders.

[Table 6.1 around here]

6.5 Empirical Results

6.5.1 Basic Herding Specification

We start with the empirical analysis by estimating the basic herding specification for full sample and two sub-samples, where CSAD is regressed only against market returns without including other effects from additional variables for each sample. Table 2 presents the results from Eq. (3.7) for the period from January 2018 to December 2020. Coefficients β_1 of all samples are positive and statistically significant at the 1% level, which indicates that the cross-sectional absolute dispersion of returns increases with the magnitude of the market returns. To assess whether herding is significant across our samples, the coefficient of interest is β_2. Chang et al. (2000) indicate the relationship between the cross-sectional absolute deviation of returns and market returns could be non-linear and the coefficient of this non-linear term is negative and statistically significant to evidence herding behaviour. In Table 6.2, this is the case for small stocks and all stocks, where β_2 is significantly negative at the 1% level and 5% level, respectively. Results suggest herding presence when traders trade on all stocks and the evidence is even stronger when traders trade on stocks with small capitalization. However, we do not find any evidence of herding for large stocks in general.

[Table 6.2 around here]

6.5.2 Effect of High Trading Intensity on Herding for Three Traders

We then test whether high trading intensity of each trader will induce herding for full sample. Table 6.3 presents results from Eq. (6.1) for three traders. If herding induces by high trading intensity, then the coefficient of the interaction term (β_4) is negative and statistically significant. According to the results, this coefficient for human traders and algorithmic traders is consistently negative but not significant, suggesting that herd behaviour is not present for either trader. However, β_4 is negative and statistically significant for market makers at the 1% level, indicating herding presence when market makers execute trades frequently.

[Table 6.3 around here]

In addition, we classify trading execution from buy-side and sell-side for each trader to analyse whether traders tend to herd when they execute buy order or when they execute sell orders. Table 6.4 presents regression results from Eq. (6.2) and Eq. (6.3) for human traders (Panel A), algorithmic traders (Panel B), and market makers (Panel C). There is no herding evidence for human traders no matter they execute buy orders or sell orders. In contract, algorithmic traders tend to herd when they execute sell orders, where coefficient of the first non-linear term is negative and statistically significant at the 5% level. But this is not because of they trade intensively, as the coefficient for the interaction term is not confirming the effect of high trading intensity on herding. For market makers, the results are consistent with results in Table 6.3. They tend to herd when they trade intensively regardless trading execution from buy-side or sell-side, as both coefficients of the interaction term (β_4) is negative and statistically significant at the 1% level.

[Table 6.4 around here]

6.5.3 Size Effect on Herding for Three Traders

McQueen et al. (1996) indicate small stocks and large stocks react fast to macro-economic news. We next test whether herding of each trader is subject to the extreme size effect. We follow Galariotis et al. (2015) to sort all stocks in our dataset according to their market capitalization on December 31st of each sample year and output the lowest 20 percent of market capitalization as small stocks, while large stocks with the top 20 percent market capitalization. Again, we identify human traders, algorithmic traders, and market makers in these sub-samples. Table 6.5 presents results of herding from Eq. (6.1) for all three traders on small stocks (Panel A) and large stocks (Panel B) respectively. Herding evidence is missing for human traders regardless of small stocks or large stocks, suggesting that human traders not tend to herd under size effect. In Panel A, algorithmic traders and market makers tend to herd in smaller capitalization stocks, evidencing by coefficients β_2 are negative and statistically significant. Moreover, β_4 is -2.98 when market makers execute trades on small stocks, this indicates herding is more pronounced when their trading intensity is high. These results are in line with previous studies which report herding evidence in smaller capitalization stocks (e.g. Lakonishok et al., 1992; Wermers, 1999; Sias, 2004; Cui et al., 2019; Andrikopoulos, Gebka and Kallinterakis, 2021).

For results of large stocks in Panel B, we find that coefficients of the interaction term (β_4) for both algorithmic traders and market makers are negative and statistically significant at the 5% level, indicating both traders have great propensity to herd in larger capitalization stocks. Large stocks with high market capitalization are mainly held and traded by institutional investors, who tend to employ algorithmic strategies. As mentioned before, market makers in this study are also known as algorithmic market makers. Therefore, the presence of significant herding for algorithmic traders and market makers can be attributed to the fact that algorithmic strategies are often have similar designs. They are not intentionally copying the same pattern, instead, they simply pursue algorithmic strategies and thus exhibit correlation in their trades. Our findings indicate that herding presence in extreme capitalization stocks, which in line with previous literature (e.g. Chang et al., 2000 and Andrikopoulos et al., 2017).

[Table 6.5 around here]

In order to examine whether traders tend to herd from one side of trading execution or from both sides of trading execution under size effect, we include dummy variables to identify high trading intensity by executing buy orders and executing sell orders respectively. Table 6.6 to Table 6.8 presents results from Eq. (6.2) and Eq. (6.3) for each trader. In Table 6.6, the presence of consistently significant and positive β_2 coefficients and β_4 coefficient for both small stocks and large stocks provide stronger support for the lack of herding on human traders. Unlike the results in Table 6.4 where algorithmic traders only tend to herd when they execute sell orders, Table 6.7 indicate the presence of herding in algorithmic traders under size effect.

More specifically, β_2 for both buy-side and sell-side for small stocks are negative and statistically significant at the 5% level, indicating algorithmic traders tend to herd when they trade on small capitalization stocks regardless of trading intensity. For large stocks, coefficients of the interaction term (β_4) from buy-side and sell-side are consistently negative and statistically significant, which provide an affirmative answer to our second hypothesis (i.e. high intensity of AT will induce herding). Algorithmic traders mainly trade on large stocks with similar algorithmic strategies. Therefore, when they trade intensively by executing more orders, the algorithms that they employ will be repeated more often (spurious herding). Small capitalization stocks would expect the relatively low trading volumes from AT, so trading intensity is not making much difference in small stocks.

[Table 6.6 around here]

[Table 6.7 around here]

Table 6.8 presents results from Eq. (6.2) and Eq. (6.3) for market makers. They tend to herd when they execute buy orders and sell order on large stocks, specifically with high trading intensity. This is evident by negative and statistically significant β_4 on both sides for large stocks. Also, the results indicate herding presence in market makers on trading execution from both-side for small stocks, where coefficients β_2 are negative and statistically significant at the 1% level. In addition, β_4 is negative and statistically significant from sell-side of small stocks, indicating market makers tend to herd when they execute intensive sell orders for small stocks.

[Table 6.8 around here]

6.6 Conclusion

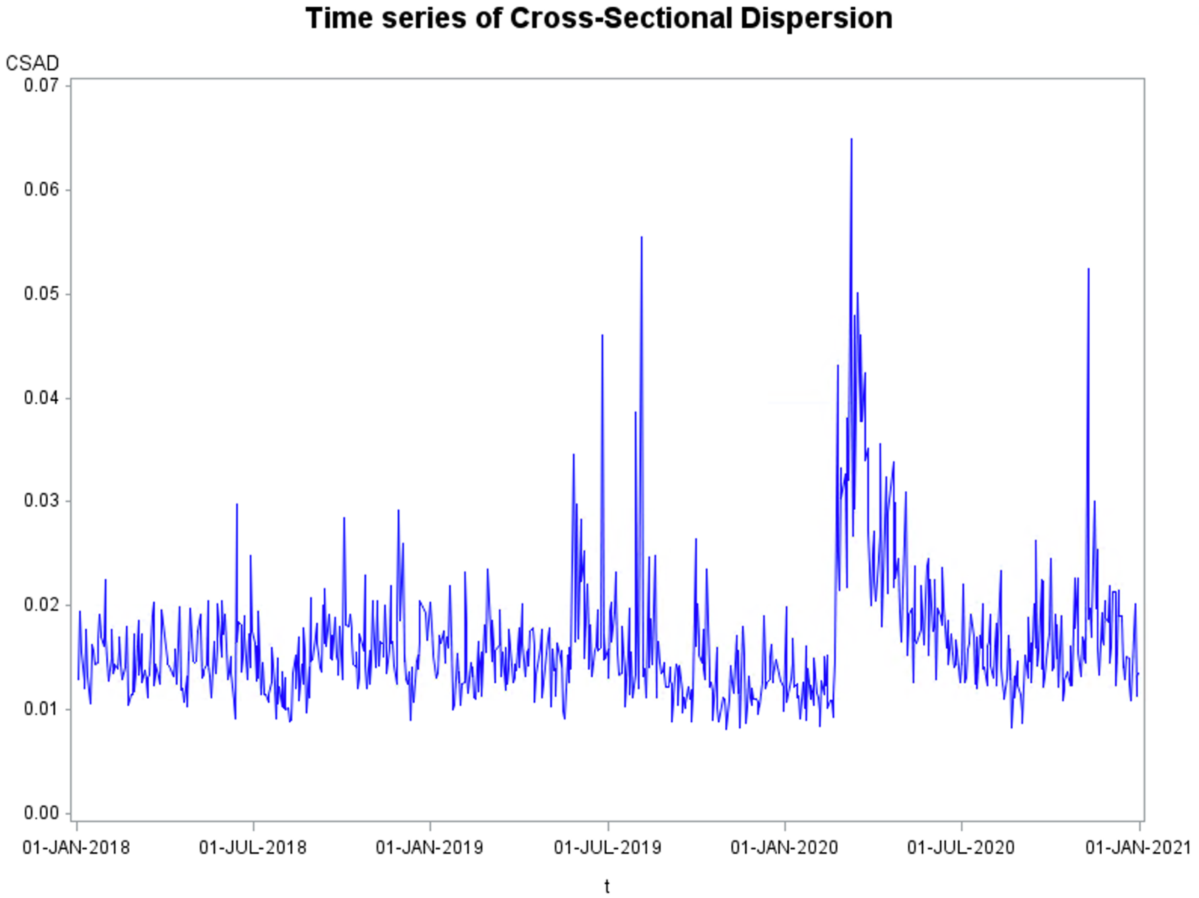
This chapter investigates for the first time in the literature of herding behaviour on three traders towards the market consensus. Original dataset contains high frequency data on Athens Stock Exchange from January 2018 to December 2020, which covering all trades and each trade can be identified from the specific trader (i.e. trade executed by human traders, or algorithmic traders, or market makers). Our first finding shows herding is significant on ASE and significant for small stocks that trade on ASE, but we find no evidence of herding for large stocks on ASE in general.

However, when we examine herding behaviour in more detail by focusing on each trader, we uncover some issues. Firstly, one of the main findings of this chapter is that traders applying algorithms induce herding behaviour, especially when the algorithm activities is intense. This is evidenced by revealing the presence of herding from algorithmic traders on extreme capitalization stocks (i.e. when they trade on small stocks and large stocks), as well as a consistent herding evidence from market makers (who using algorithm to complete trades) on full sample and sub-samples. Furthermore, when we decompose trades into buy-side and sell-side for each trader, the evidence further presents that market makers tend to herd regardless of the size of stock, and herd no matter they execute buy orders or sell orders. Similarly, algorithmic traders have great propensity to herd when they execute buy orders and execute sell orders on both small stocks and large stocks, but herding evidence only presence on the sell-side without considering the size effect.

The implication is that high intensity of trading with algorithms will induce algorithm-related herding, due to algorithm being programmed in advance. Algorithmic traders tend to trade on large stocks which normally are the most traded stocks on the exchange market, therefore, trading volumes for such stocks are normally higher than others. When algorithmic traders intensively trade on large stocks, it means the similar algorithm programme among traders will also become very active and thus induce herding, albeit it is not intentionally herding. A possible explanation for algorithmic traders only presence herding evidence on sell-side for all stocks is that algorithmic traders will flee when they suspect other more informed traders are present (e.g. market makers using algorithms). Unlike its traditional counterpart, algorithmic market makers are only on one side of the book in each stock and do not commit to provide liquidity (O'Hara, 2015). They buy when other traders are selling and they sell when other traders are buying, meaning algorithmic strategies are employing which explain the consistent herding behaviour throughout. The results are also consistent for human traders, which indicate absence of herding.

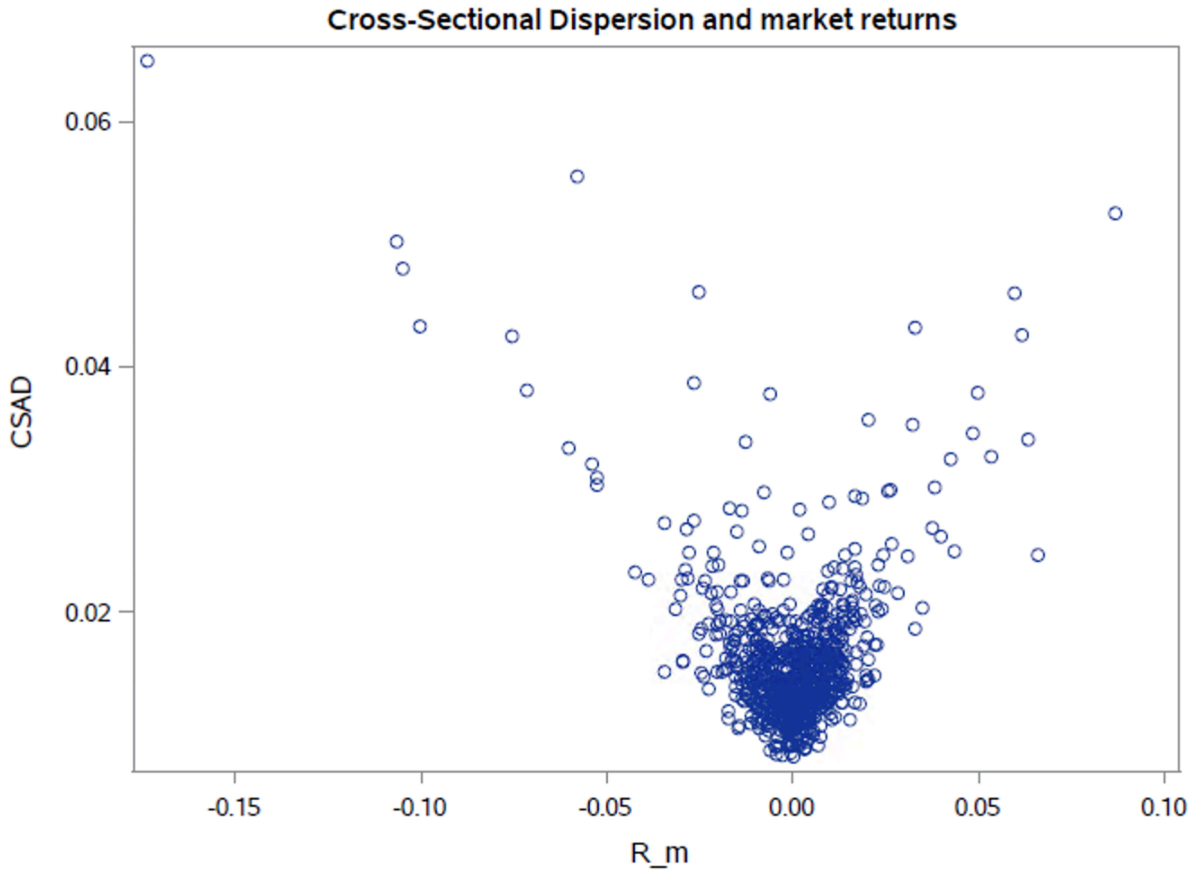
6.7 Figures and Tables

Figure 6.1



Notes: This figure plots the cross-sectional dispersion of daily stock returns between January 2018 and December 2020.

Figure 6.2



Notes: This figure plots CSAD against the daily equally-weighted market return from January 2018 to December 2020.

Table 6.1 Descriptive statistics

Panel A: Statistics for CSAD and $R_{m,t}$

	Mean		Standard deviation		Minimum		Maximum		Number of observations
	CSAD	$R_{mkt,t}$	CSAD	$R_{mkt,t}$	CSAD	$R_{mkt,t}$	CSAD	$R_{mkt,t}$	
Full sample	0.0163	-0.0004	0.0062	0.017	0.0081	-0.1737	0.0649	0.0867	744
Small stocks	0.0178	-0.001	0.0095	0.0191	0.0031	-0.1907	0.0788	0.0821	744
Large stocks	0.0141	-0.0003	0.0074	0.0206	0.0043	-0.1524	0.0611	0.1333	744

Panel B: Statistics for daily number of trades of traders

	Mean	Standard deviation	Minimum	Maximum	Number of observations
Human traders	37762.24	16161.66	9951	121479	744
Algo traders	6458.348	3921.533	856	31200	739
Market makers	6647.813	3159.046	1301	18893	744

Panel C: Statistics for daily number of trades from buy-side and sell-side

Executed buy trades					
Human traders	19262.52	8310.944	4872	61993	744
Algo traders	2737.978	1825.303	113	16299	739
Market makers	3430.405	1644.206	680	9476	744
Executed sell trades					
Human traders	18499.72	7983.464	4919	63002	744
Algo traders	3720.369	2681.56	462	26942	739
Market makers	3217.409	1549.61	621	10505	744

Notes: This table represents the descriptive statistics of the database. Panel A contains statistics on the mean, standard deviation, minimum values, maximum values, and number of observations for CSAD and $R_{mkt,t}$ for all stocks, stocks with small market capitalization, and stocks with large market capitalization. Panel B details statistics on total number of trades on daily basis of three traders. Panel C further presents statistics on trades executed on buy-side and sell-side from three traders.

Table 6.2 Baseline herding specification

	Small stocks	Large stocks	All stocks
$\lvert R_{mkt,t}\rvert$	0.4555***	0.3302***	0.3945***
	(0.0335)	(0.0264)	(0.0198)
$R_{mkt,t}^2$	-0.7754***	-0.2636	-0.4242**
	(0.2996)	(0.2578)	(0.1901)
$Adj.R^2$	0.36	0.44	0.61
Observations	744	744	744

Notes: This table presents the results for the following non-linear regression: $CSAD_t = \beta_0 + \beta_1\left|R_{mkt,t}\right| + \beta_2 R_{mkt,t}^2 + \varepsilon_t$, where CSAD is the Cross-Sectional Absolute Deviation and $R_{m,t}$ is the market return. We sort all listed stocks from our full sample (i.e. 73 stocks) each year according to the market capitalization on 31st December, then output small stocks with the lowest 20 percent of market capitalization and generate large stocks with the highest 20 percent of market capitalization. Results for the first column and the second column report the results of the herding specification for small stocks and large stocks, respectively. In the third column, we represent results of herding specification for all stocks included in our dataset. Estimations are run from January 2018 to December 2020 for each column. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 6.3 Herding on three traders

	Human traders	Algo traders	Market makers
$\lvert R_{mkt,t} \rvert$	0.3252***	0.3644***	0.3311***
	(0.0244)	(0.0232)	(0.0229)
$R^2_{mkt,t}$	0.0778	-0.1683	0.7092*
	(0.5861)	(0.3897)	(0.3878)
$Intense_{up10,t}$	0.0048***	0.0031***	0.0036***
	(0.0005)	(0.0005)	(0.0005)
$Intense_{up10,t}{R^2}_{mkt,t}$	-0.2996	-0.2019	-0.9675***
	(0.4851)	(0.3043)	(0.3051)
$Adj. R^2$	0.65	0.63	0.63
Observations	744	744	744

Notes: This table reports the results of the herding specification of full sample regarding high trading intensity of different traders from the non-linear equation $CSAD_t = \alpha + \beta_1 \lvert R_{mkt,t} \rvert + \beta_2 R^2_{mkt,t} + \beta_3 Intense_{up10,t} + \beta_4 Intense_{up10,t} R^2_{mkt,t} + \varepsilon_t$, where CSAD is the cross-sectional absolute deviation at time t, $R_{m,t}$ is the market return, $Intense_{up10,t}$ is a dummy variable that takes the value of one when the intensity of total daily trades (i.e. regardless execution of buy orders or sell orders) is located on the upper 10 percent tail of its distribution corresponding to each trader, otherwise takes zero. Estimations are run from January 2018 to December 2020 for each trader. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 6.4 Testing herding when traders executed buy and sell

Panel A: trading execution from buy-side				
	$\lvert R_{mkt,t}\rvert$	$R^2_{mkt,t}$	$Buy_{up10,t}$	$Buy_{up10,t}R^2{}_{mkt,t}$
Human traders	0.3261***	0.0701	0.0048***	-0.2945
	(0.0245)	(0.5872)	(0.0005)	(0.4859)
Algo traders	0.3645***	-0.3389	0.0029***	0.0087
	(0.0227)	(0.3458)	(0.0005)	(0.2723)
Market makers	0.3353***	0.661*	0.0035***	-0.9378***
	(0.023)	(0.3901)	(0.0005)	(0.3057)

Panel B: trading execution from sell-side				
	$\lvert R_{mkt,t}\rvert$	$R^2_{mkt,t}$	$Sell_{up10,t}$	$Sell_{up10,t}R^2{}_{mkt,t}$
Human traders	0.321***	0.0881	0.0051***	-0.2943
	(0.0243)	(0.5827)	(0.0005)	(0.4824)
Algo traders	0.3831***	-0.3648**	0.0021***	-0.0454
	(0.0202)	(0.1888)	(0.0005)	(0.2731)
Market makers	0.3282***	0.7688**	0.0037***	-1.0196***
	(0.0228)	(0.3859)	(0.0005)	(0.304)

Notes: This table shows the results of herding specification of full sample for three traders split into buy-side and sell-side. When the trader is executed buy orders, the results are given by the non-linear equation $CSAD_t = \alpha + \beta_1\lvert R_{mkt,t}\rvert + \beta_2 R^2_{mkt,t} + \beta_3 Buy_{up10,t} + \beta_4 Buy_{up10,t}R^2_{mkt,t} + \varepsilon_t$, where CSAD is the cross-sectional absolute deviation at time *t*, $R_{m,t}$ is the market return, $Buy_{up10,t}$ is a dummy variable that takes the value of one when trade execution intensity for orders from buy-side is located on the upper 10 percent tail of its distribution on daily basis, otherwise takes zero. When the trader is executed sell orders, we are testing herding specification by using $CSAD_t = \alpha + \beta_1\lvert R_{mkt,t}\rvert + \beta_2 R^2_{mkt,t} + \beta_3 Sell_{up10,t} + \beta_4 Sell_{up10,t}R^2_{mkt,t} + \varepsilon_t$, where CSAD is the cross-sectional absolute deviation at time *t*, $R_{m,t}$ is the market return, $Sell_{up10,t}$ denotes a dummy variable that takes the value of one on a day when trade execution intensity for orders from sell-side is located on the upper 10 percent tail of its distribution and zero otherwise. Estimations are run from January 2018 to December 2020 for each trader (Panels A-B). * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 6.5 Herding and the size effects on traders

	$\lvert R_{mkt,t} \rvert$	$R^2_{mkt,t}$	$Intense_{up10,t}$	$Intense_{up10,t}R^2{}_{mkt,t}$
Panel A: Herding on small stocks				
Human traders	0.4097***	-0.5494	0.0061***	-0.2246
	(0.0474)	(1.2522)	(0.0011)	(1.0405)
Algo traders	0.4468***	-0.7598**	0.0025*	0.1334
	(0.0338)	(0.3024)	(0.0013)	(0.5133)
Market makers	0.4756***	-0.8626***	0.0029***	-2.9844**
	(0.0348)	(0.3021)	(0.0011)	(1.2503)
Panel B: Herding on large stocks				
Human traders	0.2644***	-0.2374	0.0072***	0.0431
	(0.0301)	(0.5314)	(0.0007)	(0.4079)
Algo traders	0.2805***	0.3128	0.0051***	-0.6025**
	(0.0273)	(0.3416)	(0.0007)	(0.2666)
Market makers	0.2772***	0.2459	0.0064***	-0.5748**
	(0.0263)	(0.3245)	(0.0007)	(0.2571)

Notes: The results of regression $CSAD_t = \alpha + \beta_1 \lvert R_{mkt,t} \rvert + \beta_2 R^2_{mkt,t} + \beta_3 Intense_{up10,t} + \beta_4 Intense_{up10,t} R^2_{mkt,t} + \varepsilon_t$ are reported in this table. This is in effect the same regression as in Table 3 but focusing on small stocks and large stocks in the dataset. In Panel A, the results show herding specification of three traders when they trade intensively for small stocks. The results in Panel B indicate herding specification of high intensity on trading execution from each trader on large stocks. Small (Large) stocks are stocks on lower (upper) 20 percent market capitalization. Estimations are run from January 2018 to December 2020 for each size (Panels A-B). * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 6.6 Herding under high intensity trades from buy side and sell side of human traders

	Buy-side			Sell-side	
	Small stocks	Large stocks		Small stocks	Large stocks
$\lvert R_{mkt,t} \rvert$	0.3902***	0.2864***	$\lvert R_{mkt,t} \rvert$	0.4198***	0.2623***
$R^2_{mkt,t}$	0.3121	-0.5473	$R^2_{mkt,t}$	-0.8888	-0.1792
$Buy_{up10,t}$	0.0059***	0.0063***	$Sell_{up10,t}$	0.0061***	0.0071***
$Buy_{up10,t}{R^2}_{mkt,t}$	-0.9571	0.2675	$Sell_{up10,t}{R^2}_{mkt,t}$	0.0563	0.0101
$Adj. R^2$	0.39	0.50	$Adj. R^2$	0.39	0.51
Observations	744	744	Observations	744	744

Notes: This table presents results for regression $CSAD_t = \alpha + \beta_1 |R_{mkt,t}| + \beta_2 R^2_{mkt,t} + \beta_3 Buy_{up10,t} + \beta_4 Buy_{up10,t} R^2_{mkt,t} + \varepsilon_t$ in the first column (i.e. trading on small stocks) and the third column (i.e. trading on large stocks), when human traders have high execution of trades on buy-side over the sample period. The second column and the fourth column follow the regression $CSAD_t = \alpha + \beta_1 |R_{mkt,t}| + \beta_2 R^2_{mkt,t} + \beta_3 Sell_{up10,t} + \beta_4 Sell_{up10,t} R^2_{mkt,t} + \varepsilon_t$ to test herding specification when human traders specifically trade frequently on small stocks and large stocks. $Buy_{up10,t}$ takes value of one if the intensity by executing buy orders is located on upper 10 percent tail of its distribution on daily basis and zero otherwise, while $Sell_{up10,t}$ equals one if human traders execute sell orders on a day with high intensity (10 percent) of its distribution and zero otherwise. Estimations are run from January 2018 to December 2020 for each column. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 6.7 Herding under high intensity trades from buy-side and sell-side of algorithmic traders

	Buy-side			Sell-side	
	Small stocks	Large stocks		Small stocks	Large stocks
$\lvert R_{mkt,t} \rvert$	0.4518***	0.2772***	$\lvert R_{mkt,t} \rvert$	0.4487***	0.3222***
$R^2_{mkt,t}$	-0.7581**	0.4171	$R^2_{mkt,t}$	-0.7731**	-0.0926
$Buy_{up10,t}$	0.0017	0.0051***	$Sell_{up10,t}$	0.0032*	0.0027***
$Buy_{up10,t}{R^2}_{mkt,t}$	-0.0714	-0.7228***	$Sell_{up10,t}{R^2}_{mkt,t}$	0.0817	-0.7043**
$Adj. R^2$	0.36	0.48	$Adj. R^2$	0.36	0.45
Observations	744	744	Observations	744	744

Notes: This table presents results for algorithmic traders when trading execution is completed on buy-side and sell-side for small stocks and large stocks. $Buy_{up10,t}$ takes value of one if the intensity by executing buy orders is located on upper 10 percent tail of its distribution on daily basis and zero otherwise, while $Sell_{up10,t}$ equals one if human traders execute sell orders on a day with high intensity (10 percent) of its distribution and zero otherwise. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Table 6.8 Herding under high intensity trades from buy side and sell side of market makers

	Buy-side			Sell-side	
	Small stocks	Large stocks		Small stocks	Large stocks
$\lvert R_{mkt,t} \rvert$	0.4527***	0.2773***	$\lvert R_{mkt,t} \rvert$	0.4722***	0.2778***
$R^2_{mkt,t}$	-0.7441**	0.3908	$R^2_{mkt,t}$	-0.8524***	0.2364
$Buy_{up10,t}$	0.0021**	0.0059***	$Sell_{up10,t}$	0.0025**	0.0064***
$Buy_{up10,t}R^2{}_{mkt,t}$	-0.2238	-0.7484***	$Sell_{up10,t}R^2{}_{mkt,t}$	-3.0571**	-0.5628**
$Adj. R^2$	0.36	0.49	$Adj. R^2$	0.37	0.50
Observations	744	744	Observations	744	744

Notes: This table reports results when market makers execute buy-side orders and sell-side orders, corresponding to stocks with small market capitalization and stocks with large market capitalization. $Buy_{up10,t}$ takes value of one if the intensity by executing buy orders is located on upper 10 percent tail of its distribution on daily basis and zero otherwise, while $Sell_{up10,t}$ equals one if human traders execute sell orders on a day with high intensity (10 percent) of its distribution and zero otherwise. * = significance at the 10% level; ** = significance at the 5% level; *** = significance at the 1% level.

Chapter 7 Thesis Conclusion

This thesis investigates whether high frequency trading (HFT) will induce herding in equity markets and, moreover, examine herding by different traders (i.e., human traders, algorithmic traders, and market makers). To the best of my knowledge, this is the first time in the herding literature that such issues have been investigated in such a manner.

In Chapter 4, the empirical work fails to detect herding evidence in the US equity market on the aggregated daily level, this is in line with the previous studies (e.g., Spyrou, 2013). By contrast, herding is consistently present at high frequency intervals (i.e., 5-minute and 10-minute), which is probably due to the impact of increased volatility at these higher frequencies. These results evidence that HFT tends to herd while its trading intensity (proxied by number of quotes and trading volumes) is high. More specifically, the evidence suggests that HFT herds unintentionally due to similar strategies and analogous computer routes applied. This contributes to the herding literature by providing evidence of herding presence in the US equity market. It also implies that not only human-related, but also non-human trading can induce herding. Such findings provide a foundation for further research in the later chapters.

In Chapter 5, the role of HFT in equity markets is further explored, examining whether the emergence of HFT has a positive impact on the induction of herding. HFT effective date and colocation start date are employed to proxy the emergence of HFT. Because the colocation service offered by exchanges is always later than HFT effective date (when HFT starts to locate trading facilities as close as possible to the exchange, in order to take advantage of high speed by reducing the distance to the exchange), we expect the intensity of HFT activities is higher after the colocation service is available on the exchange. The empirical results show consistent and significant herding evidence by using both proxies, which indicate the emergence of HFT induces herding. Furthermore, the emergence of HFT on most exchanges in the sample are around subprime crisis period, so we also examine herding during the subprime crisis period. [42] The results indicate that herding is not triggered by the crisis, instead, the emergence of HFT induces herding during the crisis period.

In Chapter 6, the role of human traders, algorithmic traders, and market makers in herding is examined, focusing on the most traded stocks on Athens Stock Exchange from January 2018 to December 2020. The dataset commences after the implementation of MiFID II, includes all executed trades and each trade is flagged for the type of trader and whether it is an algorithmic trade. We create dummy variables to examine the effect of high trading intensity on herding from each trader. In order to test the existence of any possible size effect, we use market capitalization on 31st December to sort all stocks in the sample and generate results for small and large stocks. The findings evidence algorithmic trading (algorithmic traders and market makers) can induce herding. This implies that the high intensity of algorithmic trading induces algorithmic-related herding due to the similar algorithms programmed. Moreover, algorithmic traders and market makers tend to herd when they execute trades on small stocks and large stocks. The results are significant after we decompose trades into buy-side and sell-side of both traders. However, herding is absent for human traders. These findings contribute to the herding literature by adding evidence from non-human traders and implies that human bias is not the only factor to consider when trying to understand the generation of herding.

Overall, the thesis demonstrates the importance of HFT in herding behaviour, providing empirical evidence from different exchanges. It addresses the gap in the extent literature by bridging non-human traders (i.e. high frequency traders and algorithmic traders) and algorithm-related herding (i.e. spurious herding driven by fundamental information). Herding evidence is observed in the US equity market by investigating HFT and concludes the emergence of HFT induces herding. Moreover, we investigate herding by different traders and compare the herding

[42] It refers to the period between January 2008 and April 2011 (Galariotis et al., 2015).

effect from human and non-human traders. The findings again suggest that non-human traders induce herding on Athens Stock Exchange. In this thesis, fixed effects are included to control heterogeneity in different countries.

While we make several original contributions to this literature, several limitations could be noted. In Chapter 5, although we examine the effect of emergence of HFT on herding by using ten exchanges in nine countries, future work could consider an even wider range of exchanges and countries. In Chapter 6, regarding herding by different traders, we only access data on Athens Stock Exchange which limit its generalisability. Finally, herding appears absent by human traders in our sample, and further theoretical reasoning could address this. For example, sentiment from human perspective could be included in future research and its impact on herding considered, as well as to expand the data with more liquid exchange markets.

Finally, for further future research, this thesis recommends continuing investigation into the development of artificial intelligence, algorithms, and robotic approaches and how these may affect financial markets. There is still a lack of studies covering the impact of algorithm bias on financial exchanges, and how this differs with any potential human bias. Although HFT is following a pre-programmed algorithm, it is still designed and written by human authors. We expect to consider how sentiment transfers across from human to algorithm in the future, and the impact of decision makers in HFT firms on the outcome of algorithmic approaches.

References

Aghanya, D., Agarwal, V. and Poshakwale, S., 2020. Market in Financial Instruments Directive (MiFID), stock price informativeness and liquidity. *Journal of Banking & Finance*, 113, p.105730.

Aitken, M., Cumming, D. and Zhan, F., 2015. High frequency trading and end-of-day price dislocation. *Journal of Banking & Finance*, 59, pp.330-349.

Aitken, M., Cumming, D. and Zhan, F., 2017. Trade size, high-frequency trading, and colocation around the world. *The European Journal of Finance*, 23(7-9), pp.781-801.

Andrikopoulos, P., Gebka, B. and Kallinterakis, V., 2021. Regulatory mood-congruence and herding: Evidence from cannabis stocks. *Journal of Economic Behavior and Organization*, 185, pp.842-864.

Andrikopoulos, P., Kallinterakis, V., Leite Ferreira, M. and Verousis, T., 2017. Intraday herding on a cross-border exchange. *International Review of Financial Analysis*, 53, pp.25-36.

Bennett, J. R., Sias, R. and Starks, L., 2003. Greener pastures and the impact of dynamic institutional preferences. *Review of Financial Studies*, 16, pp.1203–1238.

Bernales, A., Verousis, T. and Voukelatos, N., 2020. Do investors follow the herd in option markets?. *Journal of Banking & Finance*, 119, p.104899.

Biais, B. and Woolley, P., 2011. High frequency trading. *Manuscript, Toulouse University, IDEI.*

Biais, B., Foucault, T. and Moinas, S., 2015. Equilibrium fast trading. *Journal of Financial Economics*, 116(2), pp.292-313.

Biffany, B. , 2018. Fixing soft dollars is not that hard: a consent and reporting framework for regulating client commission arrangement. *Duke LJ.* 68, pp.141 .

Bikhchandani, S. and Sharma, S., 2000. Herd behavior in financial markets. *IMF Staff papers*, 47(3), pp.279-310.

Black, F., 1972. Capital Market Equilibrium with Restricted Borrowing. *The Journal of Business*, 45(3), p.444.

Blasco, N., Corredor, P. and Ferreruela, S., 2011. Market sentiment: a key factor of investors' imitative behaviour. *Accounting & Finance*, 52(3), pp.663-689.

Blasco, N., Corredor, P. and Ferreruela, S., 2012. Does herding affect volatility? Implications for the Spanish stock market. *Quantitative Finance*, 12(2), pp.311-327.

Boehmer, E., Fong, K. and Wu, J., 2020. Algorithmic Trading and Market Quality: International Evidence. *Journal of Financial and Quantitative Analysis*, pp.1-30.

Boehmer, E., Li, D. and Saar, G., 2015. Correlated High-Frequency Trading. *Working Paper., Cornell University*.

Boehmer, E., Li, D. and Saar, G., 2018. The Competitive Landscape of High-Frequency Trading Firms. *The Review of Financial Studies*, 31(6), pp.2227-2276.

Boulatov, A. and George, T.J., 2013. Hidden and displayed liquidity in securities markets with informed liquidity providers. *The Review of Financial Studies*, 26(8), pp.2096-2137.

Brogaard, J., Hagströmer, B., Nordén, L. and Riordan, R., 2015. Trading fast and slow: Colocation and liquidity. *The Review of Financial Studies*, 28(12), pp.3407-3443.

Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A. and Sokolov, K., 2018. High frequency trading and extreme price movements. *Journal of Financial Economics*, 128(2), pp.253-265.

Brogaard, J., Hendershott, T. and Riordan, R., 2014. High Frequency Trading and Price Discovery. *The Review of Financial Studies*, 27, pp.2267-2306.

Cai, F., Han, S., Li, D. and Li, Y., 2018. Institutional herding and its price impact: Evidence from the corporate bond market. *Journal of Financial Economics*, 131(1), pp.139-167.

Carrion, A., 2013. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets*, 16(4), pp.680-711.

Chaboud, A., Chiquoine, B., Hjalmarsson, E. and Vega, C., 2014. Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. *The Journal of Finance*, 69(5), pp.2045-2084.

Chang, E.C., Cheng, J.W. and Khorana, A., 2000. An examination of herd behavior in equity markets: an international perspective. *Journal of Banking and Finance,* 24 (10), pp.1651–1679.

Chiang, T. and Zheng, D., 2010. An empirical analysis of herd behaviour in global stock markets? *Journal of Banking and Finance,* 34, pp.1911–1921.

Choi, N. and Skiba, H., 2015. Institutional herding in international markets. *Journal of Banking and Finance*, 55, pp.246-259.

Chordia, T., Green, T.C. and Kottimukkalur, B., 2018. Rent seeking by low-latency traders: Evidence from trading on macroeconomic announcements. *The Review of Financial Studies*, 31(12), pp.4650-4687.

Christie, W. G. and Huang, R. D., 1995. Following the pied piper: Do individual returns herd around the market? *Financial Analysts Journal,* 51, pp.31–37.

Clement, M. B. and Tse, S. Y., 2005. Financial analyst characteristics and herding behaviour in forecasting. *The Journal of Finance*, 60, pp.307–341.

Conrad, J., Wahal, S. and Xiang, J., 2015. High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics*, 116(2), pp.271-291.

Cui, Y., Gebka, B. and Kallinterakis, V., 2019. Do closed-end fund investors herd?. *Journal of Banking and Finance*, 105, pp.194-206.

Demirer, R. and Kutan, A. M., 2006. Does herding behaviour exist in Chinese stock markets? *Journal of International Financial Markets, Institutions &Money*, 16,123–142.

Dichev, I., Huang, K. and Zhou, D., 2014. The Dark Side of Trading. *Journal of Accounting, Auditing and Finance*, 29, pp.492-518.

Domowitz, I. and Yegerman, H., 2005. The Cost of Algorithmic Trading. *The Journal of Trading*, 1(1), pp.33-42.

Economou, F., Gavriilidis, K., Kallinterakis, V. and Yordanov, N., 2015. Do fund managers herd in frontier markets — and why?. *International Review of Financial Analysis*, 40, pp.76-87.

Esma.europa.eu. 2021. *MiFID II.* [online] Available at: <https://www.esma.europa.eu/policy-rules/mifid-ii-and-mifir> [Accessed 21 October 2021].

Esma.europa.eu. 2022. *MiFID II.* [online] Available at: <https://www.esma.europa.eu/policy-rules/mifid-ii-and-mifir> [Accessed 11 April 2022].

European Commission - European Commission. 2022. *Investment services and regulated markets - Markets in financial instruments directive (MiFID)*. [online] Available at: <https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-markets/securities-markets/investment-services-and-regulated-markets-markets-financial-instruments-directive-mifid_en> [Accessed 11 April 2022].

Fang, B., Hope, O., Huang, Z. and Moldovan, R., 2020. The effects of MiFID II on sell-side analysts, buy-side analysts, and firms. *Review of Accounting Studies*, 25(3), pp.855-902.

Foucault, T., 1999. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, 2(2), pp.99-134.

Foucault, T. and Menkveld, A.J., 2008. Competition for order flow and smart order routing systems. *Journal of Finance*, 63, pp.119–158.

Frino, A., Mollica, V. and Webb, R., 2014. The Impact of Co-Location of Securities Exchanges' and Traders' Computer Servers on Market Liquidity. *Journal of Futures Markets*, 34(1), pp.20-33.

Froot, K.A., Scharfstein, D.S. and Stein, J.C., 1992. Herd on the street: informational inefficiencies in a market with short-term speculation. *Journal of Finance,* 47, pp.1461–1484.

Galariotis, E., Krokida, S. and Spyrou, S.I., 2016. Bond market investor herding: Evidence from the European financial crisis. *International Review of Financial Analysis*, 48, pp.367-375.

Galariotis, E.C., Rong, W. and Spyrou, S.I., 2015. Herding on fundamental information: A comparative study. *Journal of Banking & Finance*, 50, pp.589-598.

Gavriilidis, K., Kallinterakis, V. and Montone, M., 2021. Political uncertainty and institutional herding. *Available at SSRN 3887409*.

Gavriilidis, K., Kallinterakis, V. and Ferreira, M.P.L., 2013. Institutional industry herding: intentional or spurious?. *Journal of International Financial Markets, Institutions and Money*, 26, pp.192-214.

Gębka, B. and Wohar, M., 2013. International herding: Does it differ across sectors?. *Journal of International Financial Markets, Institutions and Money*, 23, pp.55-84.

Gleason, K.C., Mathur, I. and Peterson, M.A., 2004. Analysis of intraday herding behavior among the sector ETFs. *Journal of empirical Finance*, 11(5), pp.681-694.

Guo, Y. and Mota, L., 2021. Should information be sold separately? Evidence from MiFID II. *Journal of Financial Economics*, 142(1), pp.97-126.

Hasbrouck, J. and Saar, G., 2009. Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets*, 12(2), pp.143-172.

Hendershott, T. and Moulton, P., 2011. Automation, speed, and stock market quality: The NYSE's Hybrid. *Journal of Financial Markets*, 14(4), pp.568-604.

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

Hendershott, T. and Menkveld, A.J., 2014. Price pressures. *Journal of Financial Economics*, 114(3), pp.405-423.

Hendershott, T., Jones, C. and Menkveld, A.J., 2011. Does Algorithmic Trading Improve Liquidity?. *The Journal of Finance*, 66(1), pp.1-33.

Hirschey, N., 2021. Do high-frequency traders anticipate buying and selling pressure?. *Management Science*, 67(6), pp.3321-3345.

Hirshleifer, D. and Hong Teoh, S., 2003. Herd Behaviour and Cascading in Capital Markets: a Review and Synthesis. *European Financial Management*, 9(1), pp.25-66.

Hirshleifer, D., Subrahmanyam, A. and Titman, S., 1994. Security Analysis and Trading Patterns when Some Investors Receive Information Before Others. *The Journal of Finance*, 49(5), pp.1665.

Holmes, P., Kallinterakis, V. and Ferreira, M.L., 2013. Herding in a concentrated market: a question of intent. *European Financial Management*, 19(3), pp.497-520.

Jarrow, R. and Protter, P., 2012. A Dysfunctional Role of High Frequency Trading in Electronic Markets. *Journal of Theoretical and Applied Finance*, 15, pp.219-249.

Jovanovic, B. and Menkveld, A., 2011. Middlemen in Limit-Order Markets. *SSRN Electronic Journal*.

Kirilenko, A., Kyle, A., Samadi, M. and Tuzun, T., 2017. The Flash Crash: High-Frequency Trading in an Electronic Market. *The Journal of Finance*, 72(3), pp.967-998.

Lakonishok, J., Shleifer, A. and Vishny, R., 1992. The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32(1), pp.23-43.

Malceniece, L., Malcenieks, K. and Putniņš, T.J., 2019. High frequency trading and comovement in financial markets. *Journal of Financial Economics*, 134(2), pp.381-399.

Martinez, V. and Rosu, I., 2012. High Frequency Traders, News and Volatility. *SSRN Electronic Journal*.

Scharfstein, D. S. and Stein, J. C., 1990. Herd behavior and investment. *American Economic Review*, 80, pp.465–479.

Sias, R.W., 2004. Institutional herding. *Review of Financial Studies* 17, pp.165–206.

Spyrou, S., 2013. Herding in financial markets: a review of the literature. *Review of Behavioural Finance*, 5(2), pp.175-194.

Stoll, H., 2001. Market fragmentation. *Financial Analysts Journal*, 57 (4), pp.16-21.

Theice.com. 2021. [online] Available at: <https://www.theice.com/publicdocs/IGN_Colocation_Mahwah_Operating_Policies.pdf> [Accessed 10 February 2021].

Trueman, B., 1994. Analyst forecasts and herding behavior. *Review of Financial Studies* 7, pp.97–124.

Voukelatos, N. and Verousis, T., 2019. Option-implied information and stock herding. *International Journal of Finance and Economics*, 24(4), pp.1429-1442.

Wermers, R., 1999. Mutual Fund Herding and the Impact on Stock Prices. *The Journal of Finance*, 54(2), pp.581-622.

Zhou, R. and Lai, R., 2009. Herding and information based trading. *Journal of Empirical Finance*, 16(3), pp.388-393.