The Impact of High Frequency Trading: the Nature of Informational Efficiency

Authors:
Ecaterina Bencheci
Rene Botvin

Supervisor:
Dr. Tālis Putniņš

March 2014
Riga
COPYRIGHT DECLARATION

Name of the authors in full: Ecaterina Benchefi and Rene Botvin

Title of the Thesis as approved by the SSE Riga Degree Committee

The Impact of High Frequency Trading: the Nature of Informational Efficiency

We understand that, in accordance with the Law of the Republic of Latvia regarding intellectual property rights we are the owners of the copyright in this thesis and that the copyright rests with us unless we specifically assign it to another person. We understand that the SSE Riga will deposit one copy of our thesis in the SSE Riga Library and it holds the primary right to publish our thesis or its parts in the SSE-Riga Student Research Papers series or in any other publication owned by SSE-Riga for two years from the date of our graduation. We assert our right to be identified as the authors of this thesis whenever the thesis is reproduced in full or in part.

Signed

[Signature]
Ecaterina Benchefi

[Signature]
Rene Botvin

Date
March 24th, 2014
# Table of Contents

Table of Contents ............................................................................................................. 3

1. Introduction .................................................................................................................. 4

2. Purpose of the study .................................................................................................. 6

   2.1. Research question ............................................................................................... 7

3. Literature Review ........................................................................................................ 8

   3.1. Hypotheses .......................................................................................................... 10

   3.2. Methods used in the existing literature ............................................................... 10

   3.2.1. Inefficiency .................................................................................................. 11

   3.2.2. HFT ............................................................................................................... 13

   3.2.3. Co-location ................................................................................................. 13

4. Data ............................................................................................................................ 15

5. Methodology .............................................................................................................. 17

   5.1. Measures of inefficiency .................................................................................... 17

   5.1.1. Autocorrelation ......................................................................................... 17

   5.1.2. Variance ratios ............................................................................................ 18

   5.1.3. Delay ............................................................................................................ 18

   5.1.4. Past order imbalance .................................................................................. 19

   5.2. Rolling estimation windows .............................................................................. 20

   5.3. Term structure of inefficiency metrics .............................................................. 20

   5.4. HFT measure .................................................................................................... 22

   5.5. Two stage least squares method ..................................................................... 23

6. Results ......................................................................................................................... 25

   6.1. Informational inefficiency ............................................................................... 25

   6.2. The impact of co-location ............................................................................... 26

   6.3. The relationship of HFT with other variables ................................................. 28

   6.4. The impact of HFT on informational inefficiency .......................................... 28

7. Discussion and Conclusions ...................................................................................... 31

8. References .................................................................................................................. 33

9. Appendices .................................................................................................................. 36

   9.1. Appendix A: Information about variables ...................................................... 36

   9.2. Appendix B: Inefficiency metrics .................................................................. 40

   9.3. Appendix C: Regressions results .................................................................... 45
1. Introduction

One of the most discussed topics in the field of finance at the moment is high frequency trading (HFT). This is due to the fact that HFT is a relatively new occurrence for most of the financial markets and its activity is becoming more prevalent all over the world, reaching more than half of equity trading in the United States by 2010 (Menkveld and Yueshen, 2013). HFT is a type of algorithmic trading (AT) operating based on computerised algorithms. The HFT algorithms make probabilistic predictions in the price movements of securities and depict the possibility to earn a profit from small price movements by holding a certain trading volume for a specific small amount of time.

HF traders have been defined as automated market participants that employ strategies to perform a huge amount of daily trades. HFT is differentiated from AT by its main characteristics, speed and predictive powers. It uses high-speed algorithmic programs to generate, route or execute orders within very short time windows. HF traders make very small earnings per one single stock, but given the huge amount of daily trading volume, the aggregated profits are significantly more sizable. HF traders benefit from their physical proximity to the exchange platforms, in order to minimize the latency (the ability to execute orders in time horizons not observable for regular investors) and usually end their trading day with a “flat” position (Chordia, Goyal, Lehmann and Saar, 2013).

The number of high frequency trading firms has increased over the years, with such companies as ATD, Getco, Knight Capital Group, and Citadel being present in the market already since 2006 and ATD alone trading around 800 million shares daily. Even though the profits and trading volume of such firms have decreased two fold due to the crisis, the volumes are still substantial. The HFT firms traded approximately 3.25 billion shares daily, with an average profit of a tenth of a cent per share in 2009 (Philips, 2013). Some firms were able to make profits in the amount of 4.9 billion US dollars in 2009, which then decreased to 810 million in 2012 (Mamudi and Kisling, 2014).

Over the past years, AT/HFT has become the concern of different groups of stakeholders, such as market participants, regulators and scholars. Even though many researchers claim that AT/HFT has positive effects on the measures of market quality, such as liquidity, price discovery, and volatility, the market participants do not have conclusive answers whether it might have any damaging effects in long-term horizons (Angel, Harris, and Spatt, 2010; Brogaard, Hendershott and Riordan, 2013; Hendershott, Jones, and Menkveld, 2011). Given the existing debate about this
phenomenon, the dilemma of the regulators is whether AT/HFT should be encouraged or, rather, limited.

In this paper, we take a closer look at one measure of market quality, specifically informational efficiency. We present a literature review of several studies that analyse the effects of HFT and related AT on informational efficiency in various financial markets. Our paper, however, aims to extend the existing literature by researching the nature (short versus long horizon) of the informational efficiency and the impacts of HFT on it, meaning that we intend to capture the magnitude of the HFT effects on the short-horizon and long-horizon informational efficiency and compare the two.

This paper makes an academic contribution to AT/HFT literature by adding new empirical evidence about the nature of informational efficiency that is influenced by HFT. The results are also of high interest for the market regulators since our empirical evidence makes a contribution to settling the debate regarding the nature of HFT and the need of its regulation. The market participants, especially low-frequency traders, should also appreciate our work, given that the paper seeks to capture the effect of HFT on market qualities also relevant to the regular investors.

In order to perform our analysis, we use a dataset (described in Section 4) taken from Thomson Reuters Tick History database, which provides historical trades and quotes messages for equity market with millisecond accuracy as well as full order book history of daily orders of selected trading venue. We develop our method based on several studies, such as Chordia, Roll and Subrahmanyam (2007), Comerton-Forde and Putniņš (2013), Cumming, Zhan and Aitken (2012), French and Roll (1986), Hendershott, Jones and Menkveld (2011), Hou and Moskowitz (2005), Kyle (1985), and Rösch et al. (2013). The literature review and the description of the methods in existing studies are presented in Sections 3 and 5, respectively. In Section 6 we present the empirical findings, followed by a discussion and conclusions in Section 7.
2. Purpose of the study

High frequency trading has been vastly studied by many scholars, however, there is a long list of questions that are still open and require further research. According to Chordia, Goyal, Lehmann and Saar (2013), informational efficiency, being a dimension of market quality, is affected by HFT at short intra-day horizons. This phenomenon is an important matter for the society, as the price efficiency ultimately facilitates more optimal allocation of scarce resource. It is questionable whether the millisecond environment, in which HF traders operate, can influence the welfare positively given all the expenses at which the high frequency trading takes place. This phenomenon has been studied previously, but there is still lack of empirical evidence when it comes to deeper analysis and therefore, further research is needed.

The general consensus in academic literature is that HFT improves market liquidity. Most of the studies focus on the short-horizon (millisecond) environment of the HFT and therefore it is safe to conclude, based on the existing empirical evidence, that, indeed, in very short time horizons HFT improves market quality. The nature of HFT (predicting market reactions based on past activity) however, implies that the trading does not rely on the market fundamentals. The studies that oppose HFT try to show empirical results that HFT causes increases in price volatility. If prices are more volatile and deviate from the fundamentals, they become less informative and less efficient in longer horizons. The long-horizon efficiency and accordance to fundamentals is, however, crucial for regular investors that base their valuation and investment decisions on primary financial and informational analysis. This might indicate that the short-horizon liquidity gains may not be enough to compensate for the inefficient resource allocation, which occurs due to low level of price efficiency, which can lead to the deterioration of the social welfare.

The academic literature has not yet tackled the issue regarding the impact of HFT on the market efficiency at various time horizons, leaving considerable amount of questions unanswered, which makes any new research on HFT very relevant. A fair volume of literature, which we will present in further sections, suggests that HF traders improve market efficiency. However, no empirical research has given a clear answer about the nature (short-horizon versus long-horizon) of informational efficiency influenced by HFT.
2.1. Research question

The question this research seeks to answer is following: “What is the effect of HFT on informational efficiency at different horizons? In particular, does high frequency trading improve short-horizon informational efficiency at the expense of long-horizon fundamental efficiency?”

“Short-horizon” refers to time horizons in the order of seconds to minutes, whereas “long-horizon” refers to horizons of minutes and hours to days. We measure informational inefficiency using established empirical approaches based on return predictability and deviations from random walk process.
3. Literature Review

We motivate our study based on literature that analyses the importance of informational efficiency and also papers that question the extent to which the high frequency trading impacts the market efficiency at different time horizons. Bai, Phillipon and Savov (2013) discuss the importance of the informational efficiency in the financial markets. They suggest that higher information supply leads to increased dispersion of prices, which therefore are intensively used to make income forecasts. The capital allocation can be improved when market prices provide sufficient information for investors to estimate their future earnings. Since this can be considered a measure of overall social welfare, it is important to understand at which time horizons the informational efficiency can be improved with the highest magnitude.

However, it is challenging to depict only the time variation and understand at which time intervals the efficiency improves given that the variations happen simultaneously in the markets. This is also a confrontation for Rosch, Subrahmanyam and van Dijk (2013), who studied the high frequency and low frequency market efficiency and how it varies across time and across different individual stocks. They find that the co-movement of high-frequency measures of different stocks is significant in both cross-sectional as well as time-series regressions. They also mention that answering this question from the perspective of various time horizons would benefit investors, exchange officials and policy-makers.

High frequency traders are considered to be market agents that base their trading activity on information about prices and order flow. They usually trade in opposition to price pressure. Colliard (2013) tries to show that the mispricing in short time horizons, corrected by these traders is due to speculations that might slow down the long horizons price discovery. The author analyses the behaviour of traders having an information advantage on the financial markets, as well as their impact on financial stability on the market and price discovery. Menkveld and Yueshen (2013) suggest that the benefit of information that HF traders have, positively affects the market quality, since they help to match supply with demand in short horizons and they induce market activity by giving the market participants the opportunity to learn about the true value of the assets. However, HFT may deter the market efficiency because the low-frequency traders are not able to understand whether the prices are pressured, since they cannot distinguish between inter-middlemen and middlemen-buyer trades. Hence, it is possible that the welfare is reduced in longer horizons.

Brogaard, Hendershott and Riordan (2013) also question the impact of high frequency trading on price discovery and efficiency. HF traders estimate prices based on publicly available
information and limit order book discrepancies. They do so over short time horizons of not more than 4 seconds. In order for HFT to carry out price efficiency gains, they need to demand liquidity from the other financial market participants. This imposes the question whether HF traders improve informational efficiency at the expense of the slower traders, the liquidity providers.

Research of the effects of HFT on the process of price discovery is encouraged by existing studies. For example, a recent article by Focault (2012) claims that HFT might have a negative impact on market quality and previous studies (Froot et al., 1992; Vives, 1995) bring up the possibility of informational efficiency being hindered by HFT because of its short-lived nature. Focault (2012) also mentions that HFT might trade at the expense of the low-frequency investors, who obtain the primary information that later affects the prices on the markets. Because, as mentioned before, HFT bases its trading activity mostly on secondary publicly available information, such as prices and order flows, and they do so at short horizons, the regular investors might lose motivation to attain primary information due to higher costs. This phenomenon is a barrier for the trades to be executed and low-frequency investors could exit the market. With decreased market participation, the informational efficiency of the financial markets would worsen, leading to the overall welfare decline.

The actuality of studying the impacts of HFT on market efficiency to several market participants is revealed by the number of relevant articles published recently (Bell, 2013; Clark, 2012; Dolgopolov, 2013; Zwick, 2012). The most popular topics covered include the necessity of regulation of HFT as well as the actual impact of its activity on markets. Barrales (2012) writes an article motivated by the 2010 “flash crash” and discusses about regulatory concerns related to HFT. The 2010 “flash crash” helped to draw attention towards possible drawbacks of HFT and created a considerable amount of discussions about its regulation. It was noticed that once the financial markets started to experience a downturn, the high frequency traders urged to exit their positions. At that point they did not have a market-making behaviour any more, but rather a market-taking one.

The problems rely behind market perception of HFT which does not allow regulatory powers to force high frequency traders into market-making role without other market participants considering it as unfair. The flash crash regulators had only focussed on the systematic safeguards that allowed to cut off high frequency traders under extremely volatile conditions, which would simply prevent a future equivalent flash crash. The paper concludes that regulators have not yet managed to tackle the issue properly and need to find more fundamental approaches to prevent future financial crisis.
Hence, given the unresolved puzzle in the existing research, it is questionable what the impact of HFT at different horizons is. HFT can improve market efficiency at horizons of minutes, seconds or milliseconds, though for a regular market participant it is more relevant to have efficiency over longer time horizons. Most of the academic studies claim that HFT activity in financial markets influences positively the market efficiency, although, none of these studies tested the effect of HFT on market efficiency at different, especially longer time horizons.

We are interested to observe this phenomenon and go further with our research by analysing whether the efficiency improvements might come at the expense of the efficiency at longer horizons. The topic is very relevant for many groups of market participants, as well as ultimately for the whole society through optimal resource allocation. Our paper adds value to the academic research since it brings new empirical evidence that is not present in the existing papers. It provides the readers with new insights and might raise new research questions for further studies.

3.1. Hypotheses

Derived from the background information above, we infer three hypotheses to our study:

Hypothesis 1: Short-horizon informational efficiency has improved over the past 10 years.

Hypothesis 2: Long-horizon fundamental efficiency has suffered over the past 10 years.

Hypothesis 3: HFT has significant impact on the improvement of the short-horizon informational efficiency at the expense of the long-horizon fundamental efficiency.

3.2. Methods used in the existing literature

Before moving on to the actual measures, we take a look at the Efficient Market Hypothesis (EMH), presented by Fama (1998). In his definition of EMH, he explains that a financial market is efficient when the prices reflect the available information. According to this theory, the market prices are unpredictable and follow a random walk. Since the trades are random, the activity of different types of investors is not correlated strategically and therefore they must cancel each other out without affecting the prices (Shleifer, 2000). If this was true, the activity of High frequency trading should not interfere with the activity of low frequency investors. Given that we question this phenomenon by stating our third hypothesis earlier in the paper, it is reasonable to actually bring forward measures of informational inefficiency and HFT.
Performing an empirical study involves defining the variables that we will analyse and identifying the most optimal quantitative measures for them. We are interested in the impact of HFT on informational efficiency at different horizons in time. Informational efficiency has to be measured by metrics that are valid for different time horizons (from the shortest to the longest). These measures have to be comparable across these time horizons in order to analyse the changes of the effects in the short-horizons versus long-horizons.

We select several price predictability based inefficiency measures supported by the existing literature. The previous literature has used several ways to capture the market efficiency, based on the law of one price, past order flow, intraday return predictability, and variance ratios (Rosch, Subrahmanyam and van Dijk, 2013). Many measures rely on the combination of random walk hypothesis and efficient-market hypothesis. These metrics should be computed for the listed stocks on the exchange platform at different time-horizons and across different periods of time. Such inefficiency metrics include price predictability based on stock own past returns, variance ratios, market index, past order imbalance. The ideal would be to calculate the inefficiency measures for all the stocks at very short time horizons (1sec and 10 sec), intermediate time horizons (30 sec, 60 sec, 5 min, 10 min, 30 min, 1 hour, 2 hours, 1 trading day) and very long time horizons (2 trading days and 1 trading week) for every stock-week over a period of 10 years. High frequency trading is a phenomenon that cannot be observed directly and therefore, from the econometric perspective, it has to be estimated by using a proxy.

3.2.1. Inefficiency

French and Roll (1986) investigate the difference in volatility between exchange trading hours and non-trading hours in the example of all the stocks listed on New York and American Exchange between 1963 and 1982. Their methodology proves that first-order autocorrelation can reflect the amount of information compounded in returns (and therefore also in prices), and suggests that there is a positive relation between absolute levels of the first-order autocorrelation and partial reactions or perhaps over-reactions to latest information. According to Rösch et al. (2013), the informational inefficiency metrics correlate with measures of informational inefficiency at lower frequencies. High autocorrelation implies the presence of quote deviations from a random walk and hence short-horizon predictability. This suggests that the market is highly inefficient.

Variance ratio shows how efficient and informative is the price setting of the stocks (Kyle, 1985). French and Roll (1986) suggest that greater values of variance ratios might be due to prices having incorporated more information, this being caused by a bigger number of traders possessing
private information. Hence, a higher value of such a metric suggests greater inefficiency in the market.

The Delay metric was introduced by Hou and Moskowitz (2005), who studied the significance of the impact of market frictions on individual stocks. This metric captures the extent of delayed responses of stock prices to new information. The Delay metric is described as the portion of variation in the returns of selected stocks explained by lagged market returns. They suggest that individual stocks respond to market returns and thus the market index is relevant for studying the informational inefficiency.

Other papers combine the mentioned metrics, in order to capture the effects in the most accurate way. Comerton-Forde and Putniņš (2013) take a closer look at the regulators concern about the increase in the amount of dark trading, which takes place without pre-trade transparency and should in theory harm price discovery. They use such measures as autocorrelation, variance and delay metric, based on intra-day informational inefficiencies and find that, indeed, dark trading does hinder the informational efficiency and alter the nature of price discovery. They find the harmful effect of dark trading on informational efficiency and price discovery to be present over the whole sample period and in all stocks, which provides solid underground for regulatory concerns.

Chordia, Roll and Subrahmanyam (2005) study the relationship between return predictability and order-flow. They find that the relationship is varying across time and different liquidity regimes. They find evidence that returns have been following the random-walk process better in the recent years implying that increased arbitrage activity in more liquid periods increase market efficiency. Moreover, they analyse the micro structural informational efficiency by looking at variance-based ratios and find that increased variance ratios imply more information being incorporated into prices. The evidence is consistent with the hypothesis of higher liquidity facilitating better informational efficiency. Chordia, Roll and Subrahmanyam (2005) used past-order imbalance metric as an indicator of inefficiency in their analysis of the timely information incorporation into stock prices and market efficiency.

For the high-frequency measures, Rosch, Subrahmanyam and van Dijk (2013) use popular intraday price discovery based metrics such as return predictability from order flow and past returns as well as day-to-day put-call parity abnormalities. Next they combine a market-wide measure of short-horizon efficiency by averaging across the measures of individual stocks and correlate it to market-wide variance ratios as well as to commonly used long-horizon efficiency estimates such as reversal and momentum anomalies. Their findings also suggest a positive cross-correlation between short-horizon and long-horizon measures.
3.2.2. **HFT**

HF traders place a very large amount of orders during the day given their ability of order submission at a very high speed. They base their activity on sophisticated algorithms that continuously explore trading opportunities. Hence, the increased traffic of electronic messages is a major indicator of HFT activity. The AT/HFT proxy used by Hendershott, Jones, and Menkveld (2011) relies on this intuition and has been adopted by many researchers. Their measure is calculated as the negative relative trading volume in 100 USD (relevant currency for the American market under study) to the number of electronic messages (submissions, cancellations, and trade reports) on the limit order book.

The validity of this proxy is also supported by many market participants, such as consultants Aite Group and Tabb Group, exchange platforms and other market venues. This measure was adopted by Frino and Lipone (2012), who study the relationship between the extent of market manipulation and greater occurrence of HFT. Cumming, Zhan and Aitken (2012) analyse how HFT influences the price dislocation at the end of the trading day, as well as the extent and the frequency of this occurrence. They also use the proxy for HFT across several exchange platforms. They calculate the ratio between the volume and the value of HFT trading and the same ratio at the overall market level. The HFT proxy is computed as the percentage of the first ratio out of the second one.

3.2.3. **Co-location**

Given that the markets, where HF traders operate, are distinguished by a high trading volume, with a big number of small orders and no overnight holdings, HFT are rather hard to identify (Aitken, Cumming and Zhan, 2012). Several studies have used co-location to identify HFT activity given the high correlation of the two (Boehmer, et al. 2012a, 2012b). When HF traders use co-location, it means that they rent a space next to the trading facility of the exchange platform. This allows them to profit from the time-sensitive information being transferred at an increased speed, since HF traders are connected to the exchange platform via fibre-optic cable where shorter cable connection infers a smaller latency.

Boehmer, Fong and Wu (2012a,b) used co-location as an instrument in their analysis of the impact of Algorithmic Trading on the market quality and efficiency. The authors’ motivation was that co-location events are exogenous shocks to AT that do not directly impact market quality and hence the efficiency. According to Aitken, Cumming and Zhan (2012), the exact date when co-location was introduced is not known for many exchange platforms. This is due to discrepancies in timing and uncertainty about the occurrence of HFT in a given market. The authors collected the
co-location starting dates manually from the various exchanges for several studies that they conducted (Cumming, Zhan and Aitken, 2012).
4. Data

In order to answer the research question and analyse our hypotheses, our study uses information from a low-fragmented European financial market over 10 years’ time, exploiting exogenous instrumental variables for the level of HFT, given its unobservable nature. The authors seek for a single market (not fragmented), in order to eliminate additional complications. In a fragmented market, the stocks are simultaneously traded on various trading systems in different geographical locations and therefore would impose such additional complications as decentralized trade reporting and non-synchronized timestamps from diverse places (Gresse, 2013). Most European markets are however, fragmented and rather interconnected. Moreover, it is also important that the financial market under study has exhibited noticeable changes in the HFT activity in the previous 10 years.

Swedish financial market meets all the specified requirements set above. The fragmentation of the Swedish financial market is very low, and it has experienced significant changes in the HFT activity. All the data, necessary for our study, is available on a accessible database. It is relevant for the Baltic market, as well as for other European markets. The Baltic countries have close financial linkages with the Nordic countries. Many other European markets exhibit HFT activity and the extent of its impact on these markets may be even greater than in Sweden.

In order to calculate the inefficiency metrics, we use data that includes time-stamped price-quotes of the listed stocks over the time-period of 10 years in the low fragmented market and data of the market index for the same period. For calculating the proxy of HFT, we use data that includes amount of electronic messages entered into the order book and the consolidated trading volume. We collect our data from Thomson Reuters Tick History database, which provides historical trades and quotes messages for equity markets with millisecond accuracy as well as full order book history with daily recording of all current buy and sell orders of trading venues. The data comes in the format of CSV which is easy to integrate and use (Thomson Reuters Tick History, 2012; Thomson Reuters Tick History, 2013).

Our full dataset, after formatting it through a set of filters (erroneous price quotes, incorrect ordering of time-stamps, spread outliers, etc.), contains trade and quote data consisting of 69,579,870 secondly observations per stock for the listed stocks on the Stockholm Stock Exchange. We are able to aggregate this data into the necessary frequencies in order to perform the calculations presented in our methodology. The sample period starts on 1st of January, 2004 and ends on 31st of December 2013 (the most recent date). This way we observe the changes in the
market inefficiency over a period of 10 years, before and after the beginning of HFT activity. Given the complexity of the data intensive study and the time constraints, our analysis is based on a number of stocks (selected according to the trading volume), which amounts to one fifth of the listed stocks that were included in the market index OMXS30.

We suppress 10 minutes from the start of the trading day and end of the trading day in order to eliminate unusual volatility as well as overnight returns. Hence, we end up with a consolidated order book, consisting of millisecond quotes and trades between 9:10am and 4:50pm for the selected stocks. We create control variables such as inverse midquote, spread, trading volume and number of submitted electronic messages for each individual stock using the same datasets. Electronic messages represent the sum of the modifications of the best quotes, which are identified as updates of the price or the volume of the best bid or ask.
5. Methodology

In our method, we adapt several inefficiency metrics from the studies mentioned in the previous sections in order to measure the informational efficiency at different horizons in time. By analysing how these metrics are affected by HFT, we are able to answer our research question. Due to the unobservable nature of HFT as a variable, it is necessary to identify an HFT proxy. The increased electronic message flow is an indicator of HFT activity; hence, a good proxy is the ratio between the trading volume in 1000 SEK and the number of electronic messages entered into limit order book, being consistent with existing literature. Our method includes regressions (described in further sections) using an instrumental variable that identifies the HFT activity, which, in our case, is the introduction of co-location in Sweden. The co-location facility has been introduced to the market recently and therefore suits our analysis optimally (Aitken, Cumming and Zhan, 2012). The description of all the variables used in our analysis is presented in Tables 1 and 2 (Appendix A).

5.1. Measures of inefficiency

5.1.1. Autocorrelation

We calculate the first-order return autocorrelations at various frequencies. The absolute value of the autocorrelation gives a measure of inefficiency, where higher values represent higher inefficiency. Higher absolute values of autocorrelation represent both over- and under-reaction to information of prices in the market.

We perform our calculations of first-order return autocorrelations for the selected stocks in the following manner (Comerton-Forde and Putnins, 2013):

$$ Autocorrelation_{i,t,k} = Correlation(r_{i,t,k}, r_{i,t-1,k}) $$

where $r_{i,t,k}$ is the $t^{th}$ return of stock $i$ at the frequency $k \in \{1 \text{ sec}, 10 \text{ sec}, 30 \text{ sec}, 60 \text{ sec}, 5 \text{ min}, 10 \text{ min}, 30 \text{ min}, 1 \text{ h}, 2 \text{ h}, 1 \text{ day}, 2 \text{ day}, 1 \text{ week}\}$ and is calculated based on the midquote returns as follows:

$$ r_{i,t} = 1000 \times \log \left( \frac{\text{Midquote}_{i,t}}{\text{Midquote}_{i,t-1}} \right) $$
We compute the absolute values of the autocorrelation for each stock-period. The final value suggests how the returns respond to information. As mentioned above, the smaller is the value that we get; the lower is the informational inefficiency for each specific stock-week.

5.1.2. Variance ratios

As mentioned before, one of the conditions for an efficient market is that the stock-prices follow a random walk. Variance ratios are characterized as deviations from the random walk process. If stock-prices follow a random walk, its variance of returns measured over different horizons is a linear function of the horizon, where variance of stock returns over a period of \( gt \) (where \( t \) represents a given time-period and \( g \) is a constant) is \( g \) times bigger than the variance of returns over the period of \( t \).

In order to check for market inefficiencies, we construct a simple ratio consistent with the existing empirical studies, and check whether the prices deviate from these characteristics using the formula (Comerton-Forde and Putniņš, 2013):

\[
\text{VarianceRatio}_{i,gt,k} = \left| \frac{\sigma^2_{i,gt,k}}{g \sigma^2_{i,t,k}} - 1 \right|
\]

where \( \sigma^2_{i,gt,k} \) and \( \sigma^2_{i,t,k} \) are the \( t^{th} \) and \( gt^{th} \) (\( g \) is constant) stock-week values of the returns variances of a stock \( i \) at the frequency \( k \in \{1 \ \text{sec}, 10 \ \text{sec}, 30 \ \text{sec}, 60 \ \text{sec}, 5 \ \text{min}, 10 \ \text{min}, 30 \ \text{min}, 1 \ \text{h}, 2 \ \text{h}, 1 \ \text{day}, 2 \ \text{day}, 1 \ \text{week}\} \).

5.1.3. Delay

Following the definition of an efficient market and the concept of return predictability, we introduce our next informational inefficiency metric. We examine the short-horizon and long-horizon predictability of returns based on the market index.

We construct the following regression for each stock-period (Comerton-Forde and Putniņš, 2013):

\[
r_{i,t,k} = \alpha_i + \beta_I r_{m,t,k} + \sum_{n=1}^{10} \delta_n r_{m,t-n,k} + \epsilon_{i,t}
\]

where \( r_{i,t,k} \) is the \( t^{th} \) return for stock \( i \) at the frequency \( k \in \{60 \ \text{sec}, 5 \ \text{min}, 10 \ \text{min}, 30 \ \text{min}, 1 \ \text{h}, 2 \ \text{h}, 1 \ \text{day}, 2 \ \text{day}, 1 \ \text{week}\} \) and \( r_{m,t,k} \) represents the market return, which is also lagged over 10 periods. The highest frequency at which \( \text{Delay} \) metric is calculated is 60 seconds because the data on market index returns is updated at the frequency of 60 seconds.
The \textit{Delay} metric is defined as the portion of variation in the returns of selected stocks explained by lagged market returns. It is calculated as (Comerton-Forde and Putniņš, 2013):

\[
\text{Delay}_{i,t,k} = 1 - \frac{R_i^2}{R_2^2}
\]  

(5)

where $R_i^2$ is retrieved from the results of the above mentioned regression and $R_2^2$ is retrieved from the results we have after excluding the lagged market returns from the same regression (by constraining its coefficients $\delta$ to 0). A higher value of this metric suggests higher informational inefficiency. This is due to the lagged market returns explaining a big fraction of the individual stock returns variation and the slow reaction of the stocks to the market.

5.1.4. \textit{Past order imbalance}

The fourth informational efficiency metric in our study is the predictability of returns based on past order flows. We calculate the order imbalance for each stock at different intervals in time, as follows (Rosch, Subrahmanyam and Dijk, 2013):

\[
\text{Orderimbalance}_{i,t,k} = \frac{M_{\text{buyer},i,t,k}}{M_{\text{total},i,t,k}} - \frac{M_{\text{seller},i,t,k}}{M_{\text{total},i,t,k}}
\]  

(6)

where $M_{\text{buyer},i,t,k}$ is the buyer-initiated trading volume in monetary terms, $M_{\text{seller},i,t,k}$ the seller-initiated trading volume in monetary terms and $M_{\text{total},i,t,k}$ is the total amount of money traded for stock $i$ in the stock-week $t$ at the frequency $k$ $\in \{60 \text{ sec}, 5 \text{ min}, 10 \text{ min}, 30 \text{ min}, 1 \text{ h}, 2 \text{ h}, 1 \text{ day}, 2 \text{ day}, 1 \text{ week}\}$.

The metric \text{Orderimbalance}_{i,t,k} takes into account the actual volume of trading taking place. It assigns weights to the orders according to their size. This measure allows the analysis of the change in the informational inefficiency over time, higher values indicating inefficiency.

\text{Orderimbalance}_{i,t,k} is used in the regression in the similar manner as market returns in the \textit{Delay} metric:

\[
\gamma_{i,t,k} = \alpha_i + \beta_i \text{Orderimbalance}_{i,t,k} + \sum_{n=1}^{10} \delta_i \text{Orderimbalance}_{i,t-n,k} + \varepsilon_{i,t}
\]  

(7)

where $\gamma_{i,t,k}$ is the $i^{th}$ stock-week return of a stock $i$ and \text{Orderimbalance}_{i,t,k} represents the past order flow measure for stock $i$, which is also lagged over 10 periods; all measures are taken at the
frequency \( k \in \{60 \text{ sec}, 5 \text{ min}, 10 \text{ min}, 30 \text{ min}, 1 \text{ h}, 2 \text{ h}, 1 \text{ day}, 2 \text{ days}, 1 \text{ week}\} \). The \( R^2 \) of the regressions are used to measure informational inefficiency similar to the equation 4.

### 5.2. Rolling estimation windows

For a more accurate analysis, while constructing all the inefficiency measures for every stock-week, we need enough observations within an estimation window. Hence, we use overlapping rolling estimation windows. For example, since we do not have enough daily observations in a stock-week for constructing our inefficiency metric, we extend the period of observations back several weeks so that we have in total \( N (N \geq 50) \) number of observations. As there are only 5 daily observations available per week and the estimation window for calculating the metrics is a stock-week, big part of observations (\( N - 5 \)) overlap. Overlapping rolling estimation windows are used for calculating the metrics at all horizons for assuring the consistency of the results.

### 5.3. Term structure of inefficiency metrics

The first stage of our further analysis is to unify the structure of inefficiency measures. We average all the measures such that we have only one observation for each stock per horizon per week. Having the measures of inefficiency metrics for every week over ten years, we can imagine having panel data of inefficiency „term-structures“ (where vertical axis represents inefficiency measure and horizontal axis time-horizon up to one stock-week) for every stock in question. In order to simplify the further study we make the inefficiency measures comparable by indexing all the metrics to their average across all the frequencies over the covered period. This makes the average of every indexed metric for each individual stock equal to 1. This way we can easily average all the measures of inefficiency for every stock with a simple average formula:

\[
\text{InEffMetric}_{i,t,k} = \frac{\sum_{j=1}^{4} \text{InEffMetric}_{j,i,t,k}}{4}
\]

where \( \text{InEffMetric}_{i,t,k} \) represents the average measure of inefficiency metrics \( j \in \{1,2,3,4\} \), which correspond to the inefficiency measures \{Autocorrelation\(_{i,t,k}\), VarianceRatio\(_{i,g.t,k}\), Delay\(_{i,t,k}\), Orderimbalance\(_{i,t,k}\)\}, respectively, for each stock \( i \) in the stock-week \( t \), at the frequency \( k \in \{1 \text{ sec}, 10 \text{ sec}, 30 \text{ sec}, 60 \text{ sec}, 5 \text{ min}, 10 \text{ min}, 30 \text{ min}, 1 \text{ h}, 2 \text{ h}, 1 \text{ day}, 2 \text{ day}, 1 \text{ week}\} \).
The average measures of inefficiency across stocks help us to capture market-wide changes in inefficiency. In order to make the analysis of the formed “term-structures” manageable we generate variables that summarize its shape. The variables include $\text{InEffSlope}$, $\text{InEffShort}$, $\text{InEffLong}$, $\text{InEffDifference}$. $\text{InEffSlope}$ is the “slope” of the term-structure:

$$\text{InEffSlope}_{i,t} = \text{InEffLongest}_{i,t} - \text{InEffSortest}_{i,t}$$ (9)

where $\text{InEffLongest}_{i,t}$ represents an average inefficiency measure for a stock $i$ in the week $t$ in the longest horizon (1 trading week) and $\text{InEffSortest}_{i,t}$, an average inefficiency measure for a stock $i$ in the week $t$ in the shortest horizon (1 second and 60 seconds depending on the metric).

$$\text{InEffShort}_{i,t} = \frac{\sum_{h=1\text{sec}}^{10\text{sec}} \text{InEffMetric}_{i,t,h}}{N_h}$$ (10)

where $\text{InEffShort}_{i,t}$ represents a measure of average short-horizon inefficiency metric for a stock $i$ in the stock-week $t$, $\text{InEffMetric}_{i,t,h}$ an average inefficiency measure for a stock $i$ in the stock-week $t$ at the frequency $h$ $\{1 \text{ sec}, 10 \text{ sec}, 60 \text{ sec} \}$, where $N_h$ is the total number of short time horizons.

$$\text{InEffLong}_{i,t} = \frac{\sum_{l=2\text{day}}^{5\text{day}} \text{InEffMetric}_{i,t,l}}{N_l}$$ (11)

where $\text{InEffLong}_{i,t}$ represents a measure of average long-horizon inefficiency metric for a stock $i$ in the stock-week $t$, $\text{InEffMetric}_{i,t,l}$ an average inefficiency measure for a stock $i$ in the stock-week $t$ at the frequency $l$ $\{2 \text{ days}, 5 \text{ days} \}$, where $N_l$ is the total number of long time horizons and it is equal to 2.

$$\text{InEffDifference}_{i,t} = \text{InEffLong}_{i,t} - \text{InEffShort}_{i,t}$$ (12)

where $\text{InEffDifference}_{i,t}$ represents a measure of difference between above-described average long-horizon inefficiency measure $\text{InEffLong}_{i,t}$ and average short-horizon inefficiency measure $\text{InEffShort}_{i,t}$ for a stock $i$ in the stock-week $t$. 

21
Mentioned variables that describe the “term-structure” will be used in regressions at a later stage of the methodology in order to test causal links between HFT activity and changes in inefficiency.

5.4. HFT measure

The below-described HFT proxy is an essential part of TSLS analysis. For the calculation of the HFT activity proxy we use data that captures the information about the trading volume and the number of trades and quotes for each stock co-aligning with many existing empirical studies.

As HFT servers try to minimize the latency and transaction costs, they avoid using human intermediaries and instead channel their generated orders electronically. HF traders often place and cancel orders within seconds without any order execution. From that, one can derive that the rate of electronic message traffic in an electronic limit order book is directly linked to the level of HFT activity in the market. Empirical studies that examine the linkage between market manipulation and AT/HFT activity have used the negative of a ratio between the trading volume and the number of electronic messages entered into limit order book as a proxy for the level of HFT activity. For our empirical research we use the same ratio as an approximation for the level of activity of high frequency trading in the Swedish market.

The HFT proxy calculations are executed as follows (Hendershott, Jones, and Menkveld, 2011):

\[ HFT_{proxy,i,t} = -\frac{DVol_{i,t}}{Electronic\ Msg_{i,t} \times 1000} \]  

where \( HFT_{proxy,i,t} \) represents a value of HFT proxy for stock \( i \) in the stock-week \( t \), \( DVol_{i,t} \) is the consolidated trading volume and \( Electronic\ Msg_{i,t} \) is the amount of electronic messages in the limit-order for the stock \( i \) in the given stock-week \( t \). We divide the \( DVol_{i,t} \) measure by 1000 in order to arrive at the consolidated trading volume of a specific stock for each 1000 SEK (the currency of the exchange platform under our investigation). This is consistent with Hendershott, Jones, and Menkveld (2011) who used the same ratio per 100 USD (the currency of the exchange platform under their investigation) in their study of the empirical relationship between algorithmic trading and liquidity.
5.5. Two stage least squares method

In order to be able to bring out causal relationships in our empirical analysis, we use the Two Stage Least Squares method. Since we aim to observe the impact of HFT on the informational efficiency, it is necessary to define a proxy and an exogenous instrumental variable for HFT. We have already introduced the proxy for HFT and it is a necessary measure due to the unobservable nature of HFT as a variable. Consistent with the existing studies, we use co-location as the HFT instrument in the Two Stage Least Squares analysis. It was introduced in Sweden on March 14\textsuperscript{th} of 2011 (NASDAQ OMX Nordic, 2011). It is a dummy variable being equal to 0 before co-location was introduced and 1 after its beginning. Co-location satisfies the two requirements of a valid exogenous instrumental variable: relevance (being correlated with HFT) and exogeneity (no correlation with the error term).

In the first stage of TSLS, we regress our HFT proxy on the selected instrumental variable, as follows:

\[
HFTproxy_{i,t} = \alpha_i + \gamma_t + \beta Colocation_{i,t} + \delta CV_{i,t} + \nu_{i,t}
\]  
\hfill (14)

where $\alpha_i$ represents stock $i$ fixed effects, $\gamma_t$ represents time $t$ fixed effects, and $CV_{i,t}$ is a vector of control variables, such as quoted spread, weekly trading volume, the inverse midquote ($1/\text{midquote}$) and liquidity, for each stock $i$ at time $t$. We adopt the liquidity measure from Karolyi et.al. (2012), who added a constant to the Amihud’s (2002) liquidity formula and took its logarithm in order to diminish the impact of outliers and make the measure more normally distributed. Hence, our liquidity measure is calculated as follows (Karolyi et al., 2012):

\[
Liquidity_{i,t} = -\log(1 + 1000 \ast \frac{|r_{i,t}|}{DVol_{i,t}})
\]  
\hfill (15)

where $r_{i,t}$ is the $r^\text{th}$ stock-week return of a stock $i$ and $DVol_{i,t}$ is the aggregated trading volume for stock $i$ at time $t$. The liquidity measure is also multiplied with -1 to reverse the formula and have a variable that is increasing in liquidity.

The quoted spread is expressed in bps and is calculated as follows:

\[
Spread_{i,t} = 10000 \ast \frac{\text{Ask}_{i,t} - \text{Bid}_{i,t}}{\text{Midquote}_{i,t}}
\]  
\hfill (16)
We are interested in the improved \( HFT_{proxy_{i,t}} \) for a stock \( i \) in the stock-week \( t \), which will be used in the second stage of the TSLS model.

After saving the fitted values, we start regressing on them each of the metrics that describe the shape of the inefficiency term-structure (\( \text{InEffSlope}_{i,t} \), \( \text{InEffShort}_{i,t} \), \( \text{InEffLong}_{i,t} \), \( \text{InEffDiffer}_{i,t} \)) as in the following example:

\[
\text{InEffSlope}_{i,t} = \alpha_i + \gamma_t + \beta \text{HFTproxy}_{i,t} + \delta CV_{i,t} + u_{i,t}
\]  (17)

Since we are regressing variables that describe the term-structure of inefficiency metrics on the improved value of the HFT proxy, we exclude any bias that could appear due to the part of the HFT proxy variation that could be correlated with the error term \( u_{i,t} \). Following this methodology and analysing the results, we aim to observe the impact of HFT on the informational efficiency, described by the shape characteristics of the inefficiency term-structure. Hence, we are able to answer our research question by bringing forward empirical results to support or reject our hypotheses that the efficiency in short horizons is improved by HFT at the expense of the one in long horizons.
6. Results

6.1. Informational inefficiency

In this section we analyse the resulting measures of informational inefficiency for the selected stocks from the beginning of 2005 to the end of 2013. Overall, we can notice that all four individual inefficiency metrics take higher values in long horizons when compared to short horizons. This suggests, according to our expectations, that the markets are less efficient in long time horizons than in short time horizons. When we analyse the behaviour of the average high frequency values of the inefficiency metrics over the 10 year period (see Figures 4a, 4b, 4c and 4d in Appendix B), we observe an upward trend in the Autocorrelation based informational inefficiency, downward trend in the Variance Ratio and Past Order imbalance metric and almost no trend in the Delay metric measures. Low frequency measures in general follow the trends of the respective high-frequency counterparts with the exception of being relatively more volatile over the period. For more meaningful observations we aggregate the four inefficiency measures (after indexing) and examine their overall trends at various frequencies.

Figure 5a (Appendix B) shows the shape of the aggregated informational inefficiency term structure at frequencies from 2 hours up to 5 days in any given year between 2005 and 2013. The graph shows three different stages in the transition of the term-structure over the years, which are illustrated with Figures 5c, 5b, 5d (Appendix B). In general the long-term inefficiency has been going down over the years, with the exception of the years 2009-2010. Outliers from these years are understandable considering the turbulence in the financial markets caused by world-wide financial crisis.

When we look at the changes in the inefficiency term-structure at intermediate frequencies from 60 seconds up to 2 hours, we can notice a similar trend and outlier (see Figures 6a, 6b in Appendix B). The changes in inefficiency at the high frequencies from 1 second up to 60 seconds do not display any clear outlier in any of the years (see Figures 7a, 7b in Appendix B). In this case, the transition from the year 2005 to 2013 seems to be gradual and consistent. We can also see that the decrease in short-horizon inefficiency is greater than the decrease in long-horizon inefficiency.

The slope of the term-structure at the short horizon end has been much more unpredictable and is more sensitive to external factors than the long horizon end illustrated by Figures 1a and 1b (Appendix B). The changes in the slope of the term-structure at lower frequencies are much more gradual and more persistent lasting for extended periods of time (up to a year). The slopes show
that the term-structure has flattened out over the years. One of the factors influencing the changes in the informational efficiency of the Swedish financial market could be the increased activity of HFT.

\[ \text{Slope of the inefficiency term/structure} \]

\[ \text{Slope of the inefficiency term/structure} \]

**Figure 1a. Slope of the inefficiency term-structure at 60sec-5day frequency**

Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

**Figure 1b. Slope of the inefficiency term-structure at 1sec-60sec frequency**

Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

### 6.2. The impact of co-location

We estimate the effect of the introduction of co-location in the Swedish financial market on other variables used in our paper. Table 3 (Appendix A) presents the relationship between the co-location (dummy variable) and such variables as the amount of electronic messages, the HFT proxy, quoted spread, trading volume and inverse midquote. We can observe a positive and significant relationship between the introduction of co-location and the HFT proxy, which is consistent with our expectations and the existing studies. This means that the HFT activity has increased after co-location services became available on the Stockholm Stock Exchange (see Figure 2).
We can draw a similar conclusion regarding the traffic of electronic messages. According to Table 3 (Appendix A), its coefficient denotes, ceteris paribus, an average weekly increase of approximately 15 thousand messages, for the stocks used in our analysis, after the introduction of co-location in March 2011. The average number of submitted electronic messages for these stocks before March 2011 was about 10 thousand, which suggests that the message traffic was increased more than twice with the launch of co-location (see Figure 3). Given the significant effect of the co-location on the other mentioned variables, we use them as control variables in the further analysis.
6.3. **The relationship of HFT with other variables**

Table 4 and 5 present the extent to which our HFT proxy is correlated with other variables. The relationship of HFT activity with the control variables is statistically significant at 1% significance level (Table 4). We can notice a relatively high, positive and significant correlation of the HFT proxy and the instrumental variable, co-location. This confirms that co-location can lead to an increased HFT activity. The relationship of the HFT gauge and the measure that describe the shape of the term structure are presented in Table 5. All the correlation coefficients are statistically significant at 1% significance level. The average short-horizon inefficiency is negatively correlated to the activity of HF traders, while long-horizon inefficiency is correlated positively. This suggests that an increased HFT occurrence is related to lower informational inefficiency in short time horizon and higher informational inefficiency in long time horizons. Further, we need to investigate to which extent HFT causes the informational inefficiency to change in different time horizons.

6.4. **The impact of HFT on informational inefficiency**

We perform the Two Stage Least Squares analysis in order to capture the impact of HFT activity on the informational inefficiency. Following regression 14, we obtain the fitted values for the HFT proxy from the first stage of the analysis. The regressions are run on the aggregated panel data consisting of 505 stock-week observations for each single stock over the time period between January 1st 2004 and December 31st 2013.

In the second stage of the TSLS analysis, we run twelve different models following the specification in regression 17. The dependent variables are $InEffLong_{1,t}$, $InEffShort_{1,t}$, $InEffSlope_{1,t}$, $InEffDiffe rence_{1,t}$, which account for all four inefficiency metrics, $InEffLong_{2,t}$, $InEffShort_{2,t}$, $InEffSlope_{2,t}$, $InEffDiffe rence_{2,t}$, which account for autocorrelation and variance ratio, and $InEffLong_{3,t}$, $InEffShort_{3,t}$, $InEffSlope_{3,t}$, $InEffDiffe rence_{3,t}$, which account for the Delay and Order Imbalance metrics. These are measures that describe the shape of the term structure introduced previously and all the calculations are performed according to equations 8 to 12. The regressions include the fitted HFT proxy calculated in the first stage of the TSLS analysis and a set of control variables, consisting of the quoted spread, weekly trading volume, inverse midquote and liquidity. Since these measures can impact the market efficiency, we want to use them in our models in order to capture the changes associated with them.
The results of the second stage regressions are presented in Table 6 (Appendix C). At first, we analyse the impact of HFT activity on the measures that describe the term structure of inefficiency calculated according to equation 8 and accounting for all four inefficiency metrics.

The coefficient of $\text{InEffLong}_{i,t}$ is 0.000345 and the coefficient of $\text{InEffShort}_{i,t}$ is -0.00137. We are mostly interested to compare the two coefficients, rather than examine their values separately. A relative analysis of the coefficients suggests that HF traders have a greater impact on inefficiency in short-horizons than they do in long-horizons. The direction of this impact is opposite in the two cases, being positive for long time horizons and negative for short time horizons. This means that, ceteris paribus, HFT activity improves short-horizon informational efficiency and deters the long-horizon one. Both of these effects are statistically significant at 1% significance level. The coefficient of $\text{InEffSlope}_{i,t}$ and $\text{InEffDiffe}_{i,t}$ is 0.002657 and 0.001714, respectively. Their positive sign implies that HFT activity leads to a steeper slope of the term structure of inefficiency and that the effects on long-horizon and short-horizon inefficiency happen in the same time. The impact is statistically significant at 1% significance level.

Table 6 (Appendix C) also reports the results of the regressions that account for autocorrelation and variance ratio (models 5-8) and for the Delay and Order Imbalance metrics (models 9-12). This is done in order to examine whether the outcome will display different figures because of dissimilar shortest horizons at which the inefficiency metrics are computed (due to the market index returns being reported at the highest frequency of 60 seconds).

Most of the coefficients in models 5 to 12 are statistically significant at 1% significance level. According to our results, $\text{InEffLong1}_{i,t}$ and $\text{InEffLong2}_{i,t}$ are positively affected by the increased HFT activity, only the second one being statistically significant. HFT impact on short-horizon inefficiency is negative and significant in the case of autocorrelation and variance ratio; however, it is positive and insignificant in the case of the Delay and Order Imbalance metrics. This might be due to the fact that the horizon of 60 seconds does not qualify as short-horizon, which is reasonable given the millisecond environment that HF traders operate in.

The coefficients of $\text{InEffSlope1}_{i,t}$, $\text{InEffDiffe1}_{i,t}$, $\text{InEffSlope2}_{i,t}$, $\text{InEffDiffe2}_{i,t}$ are 0.003605, 0.003031, 0.001708 and 0.000397, respectively, all of them being statistically significant at 1% significance level. The results suggest the same logic as the figures from models 3 and 4. Increased activity of HF traders influences the slope of the inefficiency term structure to become steeper, the long-horizon and short-horizon inefficiency being impacted at the same time.

We take one step further and we check the effect of high frequency trading on each of our inefficiency metrics separately. We present the results from the regressions in Table 7 (Appendix
C. The coefficients on the HFT activity measure of short-horizon autocorrelation, Delay and order imbalance metrics are all negative and statistically significant at 1% and 10% significance levels. The long-horizon autocorrelation, variance ratio and the Delay metric are influenced positively by the HFT activity and the impact is statistically significant at 1% and 5% significance levels. These additional results are consistent with our main analysis, reported in Table 6.
7. Discussion and Conclusions

Figures 5a, 6a and 7a (Appendix B) suggest that informational inefficiency diminished over the period of our study. In most cases the direction of the movements in short-horizon and long-horizon inefficiency has been the same, however the magnitude of the fluctuations are higher for inefficiency of the short horizons compared to the long horizons. Though, this is an overall overview, which comprises a rather long time window. If we take a look at the behaviour of the inefficiency term structure between the years of 2005, 2008 and 2013 (see Figures 5c, 5d and 6b in Appendix B), we will be able to argue that the short horizon inefficiency measure has experienced very little movement in this period. The long horizon inefficiency, on the other hand, has experienced a set of changes during this time, such as a considerable increase in 2009, during the times of the financial crisis.

The behaviour of the term structure of informational inefficiency measures during the year of 2009 is consistent with the existing studies, which consider HF traders to be market makers and liquidity providers during the times when the financial markets functioning well. These studies also argue that HF traders deter market quality when financial markets experience a down turn, since all the HF traders urge to exit their positions in a short period of time. The shape of the term structure for 2009 illustrates this phenomenon. It displays rather low inefficiency for the short horizons, which is beneficial for HF traders and high inefficiency for the long horizon, which is harmful for the low frequency traders.

Based on our visual analysis of the inefficiency term-structure, we cannot reject our first hypothesis that the short-horizon informational efficiency has improved over the past 10 years. We do reject our second hypothesis that the long-horizon fundamental efficiency has suffered over the past 10 years, since this is only true up until the year of 2010. The long-horizon market efficiency has improved over past three years and compensated for the significant decreases before 2010. We are not able to analyse our third hypothesis only based on the visualization tools.

According to the results of our second stage regressions, HFT activity has had a negative impact on the short-horizon and a positive impact on long-horizon inefficiency in the Swedish market over the past 10 years. Hence, we can state that the more HF traders operate on this financial market, the smaller is the short-horizon inefficiency, which is beneficial for these high frequency traders. On the other hand, the more HFT activity occurs on the financial market, the higher is the long-horizon inefficiency, which is harmful for the normal low-frequency traders. On average, HF traders impact the short-horizon informational inefficiency with a greater magnitude than the long horizon one and the fluctuations take place simultaneously. It means that HF traders
improve informational efficiency in short time horizons more than they deter the informational efficiency in long time horizons. Since this occurs simultaneously, the logic infers that the short horizon efficiency can be improved at the expense of long horizon efficiency.

Our visual analysis suggests that the inefficiency decreased both in short-horizons and long-horizons and the slope of the term-structure has become flatter. Our regression results suggest that, ceteris paribus, HFT activity increases long-horizon inefficiency. This can infer that there are more influential factors that alter the long-horizon inefficiency and overpower the effects of HFT activity. A negative impact on short-horizon inefficiency and a positive impact on long-horizon inefficiency together with several other drivers of an equal negative magnitude, would lead to an overall decrease in informational inefficiency, with the short-horizon one being greater.

This means that we cannot reject our third hypothesis that HFT has an impact on the improvement of the short-horizon informational efficiency at the expense of the long-horizon fundamental efficiency. At this point, we are able to answer our research question and conclude that high frequency trading has a significant impact on informational efficiency at various time horizons. According to the information taken from the Swedish market, HFT improves short-horizon informational efficiency at the expense of long-horizon fundamental efficiency.

Even though the deterioration of the long-horizon efficiency is much smaller than the improvement of the short-horizon efficiency, this should be a concern for market regulators. Our findings indicate that perhaps high frequency trading should be regulated at least to some degree in order to protect the low frequency traders and maintain the financial markets efficient and stable. Following sanctions and the extent of the regulation is however out of the scope of this research.

This paper should also be of interest to market participants and scholars who are curious to further research this topic. We performed an analysis using a part of the listed stocks on the Stockholm Stock Exchange and we suggest conducting this analysis on the full dataset of the listed stocks on the Stockholm Stock Exchange in order to add robustness to the results. Our findings are, nevertheless relevant for similar European markets, but can induce indication also to all the financial markets around the world. A study that would investigate this phenomenon on financial markets in other geographical regions can be based on the methodology developed in this paper and would be extremely valuable given that the performed analysis is rather complex, data intensive and time consuming. It would also fill yet another gap in the finance literature.
8. References


9. Appendices

9.1. Appendix A: Information about variables

Table 1
List of variables and their description

The table represents all the variables used for our analysis in this paper. These variables were created using panel data for the period from January 1st 2004 until December 31st 2013 for listed stocks on the Stockholm Stock Exchange. The data was extracted from Thomson Reuters database (Thomson Reuters Tick History, 2013).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{i,t,k}$</td>
<td>stock-week midqoute return at frequency $k \in {1 \ \text{sec}, \ 10 \ \text{sec}, \ 30 \ \text{sec}, \ 60 \ \text{sec}, \ 5 \ \text{min}, \ 10 \ \text{min}, \ 30 \ \text{min}, \ 1 \ \text{h}, \ 2 \ \text{h}, \ 1 \ \text{day}, \ 2 \ \text{day}, \ 1 \ \text{week}}$ measured in bps</td>
</tr>
<tr>
<td>Autocorrelation$_{i,t,k}$</td>
<td>first-order return autocorrelations at frequency $k$</td>
</tr>
<tr>
<td>VarianceRatio$_{i,t,g,k}$</td>
<td>variance of stock returns over a period $gt$ at frequency $k$</td>
</tr>
<tr>
<td>$r_{m,t,k}$</td>
<td>market index (OMXS30) return at frequency $k$ measured in bps</td>
</tr>
<tr>
<td>Delay$_{i,t,k}$</td>
<td>portion of variation in the stock returns explained by lagged market returns at frequency $k$, Delay metric</td>
</tr>
<tr>
<td>OrderimbalanceM$_{i,t,k}$</td>
<td>order imbalance in monetary terms at frequency $k$</td>
</tr>
<tr>
<td>InEffMetric$_{i,t,k}$</td>
<td>the average measure of the inefficiency metrics</td>
</tr>
<tr>
<td>InEffSlope</td>
<td>difference between average inefficiency measure in the longest horizon and average inefficiency measure in the shortest horizon (the slope of the inefficiency metrics term-structure)</td>
</tr>
<tr>
<td>InEffLongest$_{i,t}$</td>
<td>average inefficiency measure in the longest horizon</td>
</tr>
<tr>
<td>InEffSortest$_{i,t}$</td>
<td>average inefficiency measure in the shortest horizon</td>
</tr>
<tr>
<td>InEffShort$_{i,t}$</td>
<td>average short-horizons inefficiency metric</td>
</tr>
<tr>
<td>InEffLong$_{i,t}$</td>
<td>average long-horizons inefficiency metric</td>
</tr>
<tr>
<td>InEffDifference$_{i,t}$</td>
<td>difference between average long-horizons inefficiency measure and average short-horizons inefficiency measure</td>
</tr>
<tr>
<td>HFTproxy$_{i,t}$</td>
<td>negative trading volume in 1000 SEK divided by the number of electronic messages, the proxy for high frequency trading</td>
</tr>
<tr>
<td>Electronic Msg$_{i,t}$</td>
<td>Number of weekly electronic messages in the consolidated order book</td>
</tr>
</tbody>
</table>
**Table 1 - Continued**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colocation&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>a dummy variable being equal to 0 before co-location was introduced and 1 after its beginning</td>
</tr>
<tr>
<td>Liquidity&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>consolidated midquote return per 1000 SEK traded consolidated volume, Amihud liquidity measure</td>
</tr>
<tr>
<td>DVol&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>Weekly consolidated trading volume, measured in SEK</td>
</tr>
<tr>
<td>Spread&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>relative consolidated quoted spread measured in bps</td>
</tr>
<tr>
<td>InverseMidquote&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>inverse midquote for the consolidated order book, (1/midquote)</td>
</tr>
</tbody>
</table>

Source: created by the authors.

**Table 2**

**Descriptive statistics**

The table presents the descriptive statistics of selected variables from the panel data of our analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFTproxy&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>-202.24064</td>
<td>190.23394</td>
</tr>
<tr>
<td>DVol&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>2,045,881,703</td>
<td>1,497,941,928</td>
</tr>
<tr>
<td>Spread&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>19.04938</td>
<td>14.01140</td>
</tr>
<tr>
<td>InverseMidquote&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>0.00822</td>
<td>0.00450</td>
</tr>
<tr>
<td>Liquidity&lt;sub&gt;1,t&lt;/sub&gt;</td>
<td>-0.0001623</td>
<td>0.0003563</td>
</tr>
</tbody>
</table>

Source: Created by the authors using the output of data analysis in SAS.
Table 3

The impact of Co-location on other variables

The table presents the effect that the instrumental variable, Co-location, has on the other variables used in the study.

The figures in the table report the results of the following model:

$$Variable_{i,t} = \alpha_i + \gamma_t + \beta_{Colocation_{i,t}} + \varepsilon_{i,t}$$

where $Variable_{i,t}$ is one of the variables in the table below for a stock $i$ in a stock-week $t$, $Colocation_{i,t}$ is a dummy variable being equal to 0 before and 1 after co-location was introduced on the Swedish financial market, $\alpha_i$ represents stock fixed effects and $\gamma_t$ represents time fixed effects of the panel data in our analysis. The 10%, 5%, 1% significance of coefficients is indicated by *, **, ***, respectively.

<table>
<thead>
<tr>
<th>Colocation $i,t$</th>
<th>Electronic_Msg $i,t$</th>
<th>HFTproxy $i,t$</th>
<th>Spread $i,t$</th>
<th>DVol $i,t$</th>
<th>InverseMidquote $i,t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>14899.29***</td>
<td>175.9455***</td>
<td>-12.2752***</td>
<td>-7.352*(10^8)***</td>
<td>-0.00033***</td>
<td></td>
</tr>
</tbody>
</table>

Source: Created by the authors using the output of data analysis in SAS.

Table 4

The correlation between the HFT activity and selected variables

The table presents the correlation coefficients between HFT activity and the instrumental variable $Colocation_{i,t}$, as well as the control variables: $DVol_{i,t}$, $Spread_{i,t}$, $InverseMidquote_{i,t}$, and $Liquidity_{i,t}$, for each stock $i$ in a stock-week $t$. The 10%, 5%, 1% significance of coefficients is indicated by *, **, ***, respectively.

<table>
<thead>
<tr>
<th>HFTproxy $i,t$</th>
<th>Colocation $i,t$</th>
<th>DVol $i,t$</th>
<th>Spread $i,t$</th>
<th>InverseMidquote $i,t$</th>
<th>Liquidity $i,t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.49172***</td>
<td>-0.33568***</td>
<td>-0.62223***</td>
<td>-0.19453***</td>
<td>-0.12704***</td>
<td></td>
</tr>
</tbody>
</table>

Source: Created by the authors using the output of data analysis in SAS.
Table 5
The correlation between the HFT activity and inefficiency

The table presents the correlation coefficients between HFT activity and the variables that describe the term structure of the four inefficiency metrics for each stock $i$ in a stock-week $t$. The 10%, 5%, 1% significance of coefficients is indicated by *, **, ***, respectively.

<table>
<thead>
<tr>
<th>$HFTproxy_{i,t}$</th>
<th>$\text{InEffShort}_{i,t}$</th>
<th>$\text{InEffLong}_{i,t}$</th>
<th>$\text{InEffSlope}_{i,t}$</th>
<th>$\text{InEffDifference}_{i,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.11731***</td>
<td>0.27750***</td>
<td>0.13032***</td>
<td>0.23944***</td>
</tr>
</tbody>
</table>

Source: Created by the authors using the output of data analysis in SAS.
9.2. Appendix B: Inefficiency metrics

**Figure 4a. The average short and long term Autocorrelation**
(measure of informational inefficiency in the period between 2005 and 2013) Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

**Figure 4b. The average short and long term Variance Ratio**
(measure of informational inefficiency in the period between 2005 and 2013) Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).
Figure 4c. The average short and long term Delay metric
(measure of informational inefficiency in the period between 2005 and 2013) Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

Figure 4d. The average short and long term Past Order Imbalance
(measure of informational inefficiency in the period between 2005 and 2013) Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).
Figure 5a. The average 2h-5day informational inefficiency 2005-2013
Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

Figure 5b. The average 2h-5day informational inefficiency 2005-2008
Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

Figure 5c. The average 2h-5day informational inefficiency 2008-2010
Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).
Figure 5d. The average 2h-5day informational inefficiency 2010-2013
Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

Figure 6a. The average 60sec-2h informational inefficiency 2005-2013
Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

Figure 6b. The average 60sec-2h informational inefficiency 2005-2013 transition
Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).
Figure 7a. The average 1sec-60sec informational inefficiency 2005-2013
(includes measures of autocorrelation and variance ratio) Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).

Figure 7b. The average 1sec-60sec informational inefficiency 2005-2013 transition
(includes measures of autocorrelation and variance ratio) Source: created by the authors using data from Thomson Reuters database (Thomson Reuters Tick History, 2013).
9.3. Appendix C: Regressions results

Table 6

HFT impact of informational inefficiency

This table presents the results from regressions in the second stage of the TSLS model. The dependent variables from the regressions were run on the fitted HFT proxy and a set of control variables. The 10%, 5%, 1% significance of coefficients is indicated by *, **, ***, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dependent Variable</th>
<th>( HFT_{proxy,i,t} )</th>
<th>( DVol_{i,t} )</th>
<th>( Spread_{i,t} )</th>
<th>( InverseMidquote_{i,t} )</th>
<th>( Liquidity_{i,t} )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>( InEffLong_{i,t} )</td>
<td>0.000345***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>10.40455***</td>
<td>134.3055***</td>
<td>0.3741</td>
</tr>
<tr>
<td>2.</td>
<td>( InEffShort_{i,t} )</td>
<td>-0.00137***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>25.80803***</td>
<td>-2.49122</td>
<td>0.4463</td>
</tr>
<tr>
<td>3.</td>
<td>( InEffSlope_{i,t} )</td>
<td>0.002657***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>4.497357</td>
<td>78.21138</td>
<td>0.5188</td>
</tr>
<tr>
<td>4.</td>
<td>( InEffDifference_{i,t} )</td>
<td>0.001714***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>-15.4035***</td>
<td>136.7967**</td>
<td>0.3969</td>
</tr>
<tr>
<td>5.</td>
<td>( InEffLong1_{i,t} )</td>
<td>0.000114</td>
<td>0.00***</td>
<td>0.00***</td>
<td>24.81696***</td>
<td>253.075***</td>
<td>0.4340</td>
</tr>
<tr>
<td>6.</td>
<td>( InEffShort1_{i,t} )</td>
<td>-0.00292***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>32.57431***</td>
<td>67.5511</td>
<td>0.5060</td>
</tr>
<tr>
<td>7.</td>
<td>( InEffSlope1_{i,t} )</td>
<td>0.003605***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>49.85021***</td>
<td>236.0858*</td>
<td>0.4809</td>
</tr>
<tr>
<td>8.</td>
<td>( InEffDifferencel_{i,t} )</td>
<td>0.003031***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>-7.75735</td>
<td>185.5239</td>
<td>0.4970</td>
</tr>
<tr>
<td>9.</td>
<td>( InEffLong2_{i,t} )</td>
<td>0.000577***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>-4.00785*</td>
<td>15.53596</td>
<td>0.5642</td>
</tr>
<tr>
<td>10.</td>
<td>( InEffShort2_{i,t} )</td>
<td>0.00018</td>
<td>0.00***</td>
<td>0.00***</td>
<td>19.04174***</td>
<td>-72.5335*</td>
<td>0.6827</td>
</tr>
<tr>
<td>11.</td>
<td>( InEffSlope2_{i,t} )</td>
<td>0.001078***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>-40.8555***</td>
<td>-79.663</td>
<td>0.8029</td>
</tr>
<tr>
<td>12.</td>
<td>( InEffDifferencel2_{i,t} )</td>
<td>0.000397***</td>
<td>0.00***</td>
<td>0.00***</td>
<td>-23.0496***</td>
<td>88.0695*</td>
<td>0.6729</td>
</tr>
</tbody>
</table>

Source: Created by the authors using the output of data analysis in SAS.
This table presents the coefficients resulting from regressions in the second stage of the TSLS model. The dependent variables (longest and shortest horizons inefficiency metrics) from the regressions were run on the fitted HFT proxy and a set of control variables. The 10%, 5%, 1% significance of coefficients is indicated by *, **, ***, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dependent Variable</th>
<th>( HFT_{\text{proxy},i,t} )</th>
<th>( DVol_{i,t} )</th>
<th>( \text{Spread}_{i,t} )</th>
<th>( \text{InverseMidquote}_{i,t} )</th>
<th>( \text{Liquidity}_{i,t} )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>AutocorrLong_{i,t}</td>
<td>0.000135***</td>
<td>0.00***</td>
<td>0.002***</td>
<td>0.00***</td>
<td>13.894***</td>
<td>0.3586</td>
</tr>
<tr>
<td>2.</td>
<td>AutocorrShort_{i,t}</td>
<td>-0.00011***</td>
<td>0.00***</td>
<td>0.00042***</td>
<td>0.00***</td>
<td>-6.13287***</td>
<td>0.6621</td>
</tr>
<tr>
<td>3.</td>
<td>VarianceRLong_{i,t}</td>
<td>0.00173***</td>
<td>0.00***</td>
<td>0.0329***</td>
<td>0.00***</td>
<td>110.504</td>
<td>0.2951</td>
</tr>
<tr>
<td>4.</td>
<td>VarianceRShort_{i,t}</td>
<td>0.001602***</td>
<td>0.00***</td>
<td>0.0575***</td>
<td>0.00***</td>
<td>286.2747**</td>
<td>0.3984</td>
</tr>
<tr>
<td>5.</td>
<td>DelayLong_{i,t}</td>
<td>0.000802***</td>
<td>0.00***</td>
<td>0.0096***</td>
<td>0.00***</td>
<td>102.0793***</td>
<td>0.8449</td>
</tr>
<tr>
<td>6.</td>
<td>DelayShort_{i,t}</td>
<td>-0.00019**</td>
<td>0.00***</td>
<td>0.0052***</td>
<td>0.00***</td>
<td>-0.24259</td>
<td>0.8315</td>
</tr>
<tr>
<td>7.</td>
<td>OrderimbLong_{i,t}</td>
<td>0.00</td>
<td>0.00***</td>
<td>0.0003</td>
<td>0.00***</td>
<td>-4.19174</td>
<td>0.8511</td>
</tr>
<tr>
<td>8.</td>
<td>OrderimbShort_{i,t}</td>
<td>-0.00015***</td>
<td>0.00***</td>
<td>0.0049***</td>
<td>0.00***</td>
<td>-34.6185**</td>
<td>0.8444</td>
</tr>
</tbody>
</table>

Source: Created by the authors using the output of data analysis in SAS.