Economic Report

High-frequency trading activity in EU equity markets

Number 1, 2014

Table of Contents

Executive summary

The objective of this report is to shed further light on the extent of high-frequency trading (HFT) in EU equity markets. We use unique data collected by ESMA, covering a sample of 100 stocks from nine EU countries for May 2013. Our study complements the HFT literature by looking at equity markets across a number of EU countries. Most of the HFT studies published so far focus either on the US or on a single country within Europe.

One of the challenges faced by empirical studies is the operational definition of HFT. There is a variety of approaches in the literature to estimate HFT activity. None of these approaches is able to exactly capture HFT activities and they lead to widely differing levels of HFT activity. This is an important issue for the analysis of HFT activity and its impacts. It is also a significant challenge for regulators who need to define what constitutes HFT activity. The approach taken in this report is to provide a lower and an upper bound for HFT activity.

Two main approaches have been used in the literature: i) a direct approach based on the identification of HFT firms according to their primary business or the types of algorithms they use, and ii) an indirect approach based on statistics such as lifetime of orders or order-to-trade ratio.

We provide estimations for HFT activity based on the primary business of firms (direct approach) and based on the lifetime of orders (indirect approach). The first proxy is an institution-based measure (each institution is either HFT or not), while the second proxy is a stock-based measure (an institution may be HFT for one stock but not for another one).

The results based on the primary business of firms provide a lower bound for HFT activity, as they do not capture HFT activity by investment banks, whereas the results based on the lifetime of orders are likely to be an upper bound for HFT activity.

In our sample, we observe that HFT activity accounts for 24% of value traded for the HFT flag approach and 43% for the lifetime of orders approach. For the number of trades the corresponding numbers for HFT activity are 30% and 49%, and for the number of orders 58% and 76%. The difference in the results is mainly explained by HFT activity of investment banks which is captured under a lifetime of orders approach, but not under a HFT flag approach (see table C.1).

HFT activity - overall results for the HFT flag and lifetime of orders approaches					C.1
	HFT flag	Lifetime of orders			
		lifetime of orders: total	thereof HFT firms	thereof investment banks	thereof other firms
Value traded	24	43	19	22	2
Number of trades	30	49	25	23	2
Number of orders	58	76	55	19	1

Note: Figures are weighted by value of trades (value traded), number of trades, and number of orders; in %.
Source: ESMA.

Our results also show that the level of HFT activity varies widely between trading venues. We also observe that HFT activity is linked to market capitalisation with HFT activity increasing with the market capitalisation of stocks.

This report describes the results of the first part of the ESMA research on HFT. Further research is needed regarding

— the drivers of HFT activity,

— to assess the actual contribution of HFT to liquidity, and

— to analyse potential risks and benefits linked to HFT activity.

High-frequency trading activity in the EU

Introduction

Over the last few years, financial markets have undergone a series of significant changes. Regulatory developments, technological innovation and growing competition have increased the opportunities to employ innovative infrastructures and trading practices. On the regulatory side, the entry into force of the Market in Financial Instruments Directive (MiFID) in 2007 has re-shaped markets in the EU. At the same time, developments in trading technologies enabled the use of automated and very fast trading technologies. The resulting trading landscape can be characterised by higher competition between trading venues, the fragmentation of trading in the same financial instruments across venues in the EU as well as the increased use of fast and automated trading technologies.

These developments have interacted with each other. On the one hand, increases in competition as well as in the dispersion of trading may have boosted the use of algorithmic trading. On the other hand, increased competition could have been possible, at least partly, because of high-frequency trading (HFT) activity, as HFT is able to integrate activity on different venues.[1]

At the same time, a series of events such as the May 2010 Flash Crash in the US, problems faced during BATS and Facebook IPOs and the loss of USD 420mn by Knight Capital in August 2012 due to a malfunctioning algorithm have called into question the benefits and risks linked to algorithmic and high-frequency trading. In particular, the impact of HFT on volatility, liquidity and, more generally, market quality has been an important topic for securities market regulators, academics and market practitioners.

While the academic and policy-oriented literature had originally focused predominantly on US markets, research into EU equity markets has been increasing in recent years. The focus typically is on a specific EU country and/or trading venue.[2]

The objective of this report is to shed further light on the extent of HFT on EU equity markets using unique data collected by ESMA. In particular, this report discusses the identification of HFT and provides estimates of HFT activity based on a cross-EU sample of stocks.

The structure of the paper is as follows. First, an overview of different methods for identifying HFT activity is provided. Second, we describe our dataset, after which indications for the extent of HFT activity in EU equity markets are provided. The last section concludes.

Definition and identification of high-frequency trading activity

Algorithmic trading (AT) and high-frequency trading (HFT) are trading practices which are still relatively recent and which are still evolving. A precise definition is thus emerging only slowly. A legal definition is provided by MiFID II.[3] In a research context, the academic literature is narrowing down its definition to a few identifying features.

In general, total trading activity can be divided into algorithmic trading (AT) and non-algorithmic trading, depending on whether or not market participants use algorithms to make trading decisions without human intervention. Kirilenko and Lo (2013), for example, describe AT as "the use of mathematical models, computers, and telecommunications networks to automate the buying and selling of financial securities".

Brogaard (2012) describes high frequency traders, in turn, as "the subset of algorithmic traders that most rapidly turn over their stock positions." Following definitions proposed in the literature, HFT has the following features

— proprietary trading;

— very short holding periods;

— submission of a large number of orders that are cancelled shortly after submission;

— neutral positions at the end of a trading day; and

— use of colocation and proximity services to minimise latency.[4]

From an analytical perspective, the absence of a unique definition makes it difficult to achieve a precise identification of HFT activity. The literature employs a number of approaches to identify HFT activity. These use one or several of the aforementioned characteristics and lead to differing results in assessing the level of HFT activity in equity markets.

---

[1] See Pagano (1989) for the general argument and Biais and Woolley (2011) for the application in the context of HFT.

[2] See Annex 1 for a survey of the literature on HFT identification. For a survey of the HFT literature see also U.S. Securities and Exchange Commission (2014). Specific examples for studies of European markets are inter alia: Brogaard et al. (2014) for LSE, Gomber and Gsell (2009) and Hendershott and Riordan (2013) for Xetra, Hagströmer and Nordén (2013) for Nasdaq-OMX, Jovanovic and Menkveld (2010) for NYSE Euronext Amsterdam and Chi-X.

[3] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0065&from=EN
Article 4(1)(39) of MiFID II states that algorithmic trading "means trading in financial instruments where a computer algorithm automatically determines individual parameters of orders such as whether to initiate the order, the timing, price or quantity of the order or how to manage the order after its submission, with limited or no human intervention, and does not include any system that is only used for the purpose of routing orders to one or more trading venues or for the processing of orders involving no determination of any trading parameters or for the confirmation of orders or the post-trade processing of executed transactions".

Article 4(1)(40) of MiFID II describes a high-frequency trading technique as "an algorithmic trading technique characterised by: (a) infrastructure intended to minimise network and other types of latencies, including at least one of the following facilities for algorithmic order entry: co-location, proximity hosting or high-speed direct electronic access; (b) system-determination of order initiation, generation, routing or execution without human intervention for individual trades or orders; and (c) high message intraday rates which constitute orders, quotes or cancellations".

[4] See U.S. Securities and Exchange Commission (2010).

In the following we describe different approaches used in the literature. These fall into two broad categories – direct and indirect approaches.

We focus on methodological advantages and disadvantages both in a general context and in the context of our dataset.

Direct approach

The direct approach to identify HFT activity relies on the identification of market participants either based on

- their primary business and / or
- the use of services to minimise latency.

The former method is used by Brogaard et al. (2013a) and Brogaard et al. (2013b). Information on the primary business of firms is either obtained by the trading venue and/or by the authors of the study.[5] This approach focuses on pure HFT firms which are flagged as HFT firms. It does not cover HFT activity by other firms, such as HFT activity carried out by investment banks. It may not include activity by HFT firms routing their trading activity through another trading venue member (direct market access or sponsored access), unless the broker reports the HFT firms as clients. The HFT flag approach also implies that all trading by the identified firms is considered HFT, while in practice they may use HFT and non-HFT strategies.

Consequently, relying on the primary business of firms has an element of underestimation (HFT activity by other non-HFT firms not counted) as well as an element of overestimation (not all activity by HFT firms is in fact HFT). Firms with HFT as primary business will in all likelihood predominantly use HFT strategies. Therefore it is likely that the underestimation element is dominating. This is corroborated by the analysis carried out with our data set. The HFT flag approach provides the lowest estimates for HFT activity. Additionally, under the lifetime of orders approach, most of the trading activity carried out by HFT firms is identified as HFT activity. Therefore, we consider the HFT flag approach to provide for a lower bound in terms of estimation of HFT activity.

The second direct method to identify HFT activity relies on prior information on the use of low-latency infrastructure - e.g. the use of colocation and proximity services or access to fast data feeds. This approach does not require any knowledge of the firm's primary business, but can be too encompassing as brokers trading exclusively on behalf of their clients (agent trading) may also use colocation services to offer best execution strategies to their clients. Therefore, relying only on colocation would inflate statistics on HFT activity. This is corroborated by our analysis.

One possibility to deal with the overestimation issue would be to focus on proprietary trading by participants using colocation services. However, this is difficult in practice, as flags for proprietary and agent trading may not to be fully consistent across venues, making a cross-country and cross-venue comparison difficult.

Indirect approach

Indirect approaches rely on the trading and quoting patterns of market participants. Identification based on inventory management, for example, is closely related to the broader concept of trading patterns. Examples related to quoting patterns are identification based on the lifetime of orders, message traffic, order-to-trade ratios and HFT firm strategies.

1) Intraday inventory management

Using trade data, Jovanovic and Menkveld (2012) and Kirilenko et al. (2010) define HFT firms as intermediaries with high volumes traded and low intraday and overnight inventories, in line with the main characteristics of HFT. Chart C.2 shows an illustrative example based on mock-up data. Member 2 is a net seller of the stock, while Members 1 and 3 manage their inventories so to have a flat position at the end of the day. However, Member 3 manages its inventory intraday in order to have a flat position also during the trading day, while Member 1 has a net seller position throughout most of the trading day. In this example, Member 3 would be flagged as HFT.

Intraday inventory management   C.2

Note: Net inventory of firms, based on mock-up data.
Source: ESMA.
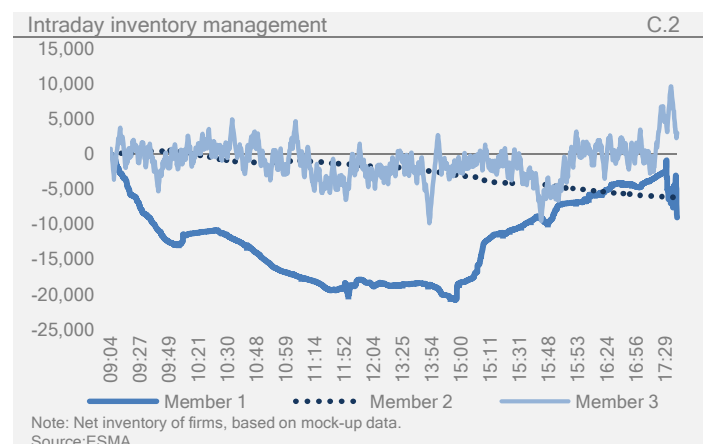
Identification based on intraday inventory management will tend to identify high-frequency market making strategies and may not identify other strategies.

In addition, to have a complete picture of intraday inventory management of firms, data on equity are likely to be insufficient as positions in related financial instruments such as equity swaps, ETFs or equity futures will not be considered.

2) Lifetime of orders

An alternative approach consists in looking at the lifetime of orders, i.e. the time elapsed before the order is modified or cancelled. Chart C.3 is based on our dataset. It shows the lifetime of orders for HFT firms, investment banks and other firms. Firms identified as HFT under the direct approach appear to send orders with shorter lifetime (40% less than 0.2 seconds), compared to Investment banks (40% less than 5 seconds) and other firms (40% less than 3 seconds).

---

[5] Brogaard et al. (2013b) used an additional list based on the identification of type of participants by trading venues and based on supervisory knowledge.

Distribution of lifetime of orders   C.3

Note: Percentiles of the lifetime of cancelled or modified orders, in seconds
Source: ESMA
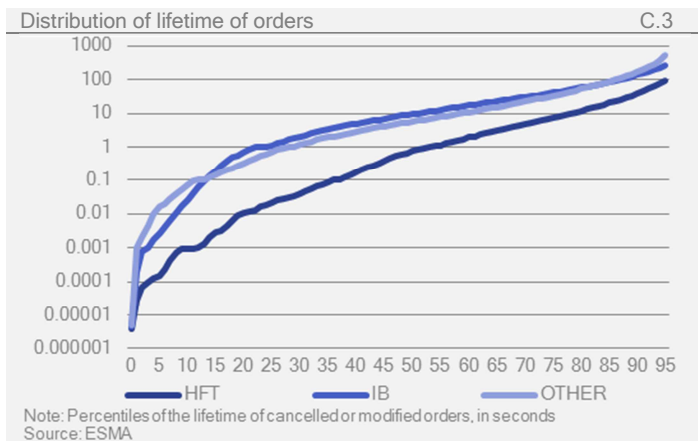
However, some firms, not flagged as HFT under the direct approach, may also be able to send very quick orders. Chart C.4 indicates that the 25% quickest investment banks and the 25% quickest other firms (first quartile) were able to modify or cancel their orders faster than the median of HFT firms.

Distribution of lifetime of orders   C.4

Note: Percentiles of the lifetime of cancelled or modified orders of all HFT firms and first quartile for investment banks and other firms, in seconds.
Source: ESMA
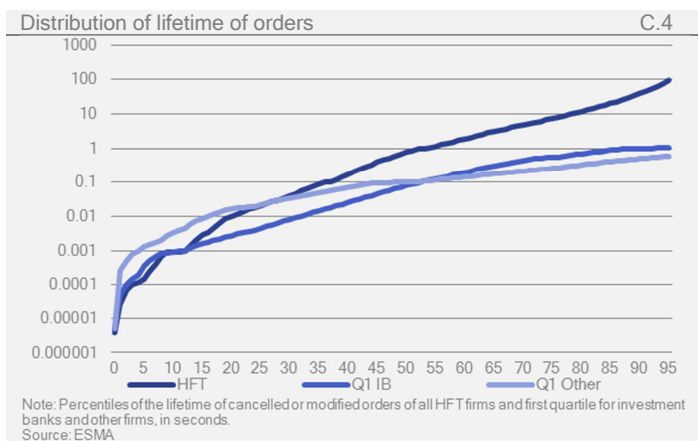
Hasbrouck and Saar introduce two identification approaches related to the lifetime of orders. The concept of "fleeting orders" originates from Hasbrouck and Saar (2009) and is defined as an order that is added and removed from the order book within a given short period of time (x milliseconds). Hasbrouck and Saar (2013) introduce the concept of strategic runs, which they define as a sequence of linked order book messages. More precisely, strategic runs are defined as a series of at least 10 submissions, cancellations and executions, sent consecutively in less than one second over a 10-minute interval. The classification of HFT activity according to the lifetime of orders is based on the ability of a market participant to very quickly modify or cancel orders and can be computed for individual stocks rather than at firm level. Firms may have HFT activity in some stocks, but not in others. The lifetime approach can identify trading activity in stocks where firms act as HFT.

To operationalise this approach several dimensions need to be considered. Firstly, order lifetime thresholds to classify trading activity as HFT or non-HFT need to be set. Secondly, the level of granularity for the analysis regarding the trading activity of firms needs to be defined.

Order lifetime thresholds to identify HFT activity can be set in a number of ways and have both absolute and relative dimensions.

One possibility is to set thresholds both in an absolute dimension (x milliseconds) and in a relative dimension (y% of orders need to be modified in less than x milliseconds). The absolute dimension captures the speed of trading activity; the relative dimension takes into account that some firms might use 'slow' orders alongside HFT activity (e.g. investment banks for agent orders). This type of approach is well suited to datasets which identify HFT activity at a certain point in time, but does not take into account that trading speed changes over time through technological progress.

An alternative is using purely relative thresholds. Here activity would be classified as HFT when the lifetime of orders of a firm is lower than e.g. the median of the lifetime of all orders on a trading venue. This threshold can be set at the level of an individual stock or for all trading activity of a firm on a trading venue. Setting purely relative thresholds takes account of changes in the speed of overall trading activity, however, the relative threshold needs to be calibrated carefully.

For both the absolute and the relative approach, there is no rule which threshold would characterise HFT activity in a precise manner. Therefore thresholds need to be calibrated carefully and robustness checks should be carried out.

For both approaches, there is a potential to both under- and overestimate HFT activity. Trading activity of a group is either classified as HFT or not. HFT activity might be underestimated if only a small portion of a group's trading activity involves very quick order cancellations and modifications. On the other hand, there is also a distinct possibility of overestimating HFT activity. This would particularly be the case for investment banks, who are captured as HFT based on their overall trading patterns, although HFT is likely to constitute only part of their total trading activity. Hence, these approaches could capture both the activity of an HFT desk and the slower agent-based trading activity. On balance, it is likely that the lifetime of orders approach at a group level overestimates HFT activity

Our dataset provides a snapshot of trading activity over a short time period. We therefore use absolute thresholds for our headline results[6]. Regarding the aforementioned overestimation issues, there are two ways to mitigate these. Firstly, a flag in the data for proprietary and agent orders could be used as a proxy. While our dataset has such a flag, the quality of the data is insufficient to use it.

Secondly, the lifetime of orders can also be calculated at the firm or participant level, instead of at group level. This would mean that only the activity by "fast" participants within a group would be captured as HFT activity, whereas

---

[6] For reference we also describe how much trading activity would be classified as HFT activity under a purely relative threshold in Annex 2.

the activity of the "slower" firms within the group would not. This approach is feasible in our dataset. However, data issues were encountered, leading to biases in the results (see Annex 2 for the results and a description of the data issues). Therefore, we calculate the lifetime of orders at group level.

3) Message traffic (including order-to-trade ratios)

Proxies based on message traffic have been used to identify AT and HFT by academics, industry bodies, trading venues and regulators. Hendershott et al. (2011) use the number of messages per $100 of trading volumes along with message-to-trade ratios. The German HFT Act, which came into force on 15 May 2013, uses message traffic as one of the elements to identify HFT firms. Firms generating message traffic of more than two messages per second or 75,000 messages per trading day are considered to be HFT firms if they also fulfil the other criteria of the HFT act. These relate to the use of infrastructure intended to minimise latency and system determination of individual orders and trades, i.e. orders are initiated, generated, routed or executed without human intervention.
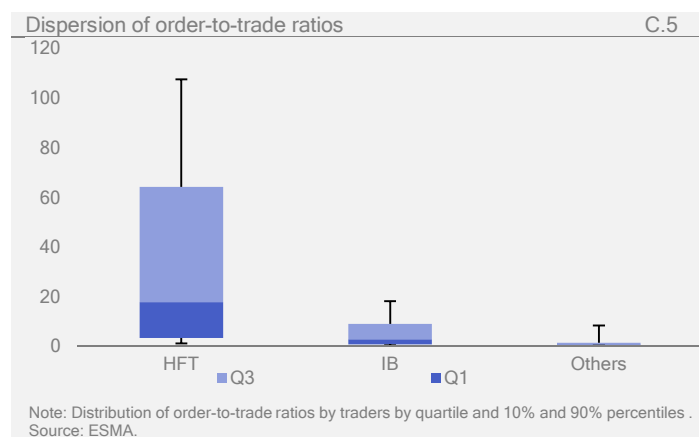
In North America, a report on HFT published by the Investment Industry Regulatory Organization of Canada (2012) uses order-to-trade ratios (OTR) as a proxy for HFT. High order-to-trade ratio ("HOT") traders were identified as the ones with the largest ratios compared to the entire sample.[7] The Australian Securities & Investments Commission (2013) uses a mix of indicators to identify HFT in Australia, one of them being OTR.

However, using only OTR may lead to biases in the results. Firstly, this approach identifies mostly passive HFT strategies, such as market making where participants regularly update their bid and ask quotes, resulting in high OTR.[8] Statistical arbitrage strategies, which rely on low latency, would not be captured by this metric as they do not require high OTR. Secondly, algorithms used by firms for agency trading on behalf of institutional investors may result in high OTR and therefore be mislabelled HFT, as noted by Malinova, Park and Riordan (2013). Thirdly, some firms may only have executed few trades (or none at all) despite having sent orders, resulting in very high OTR. This might be particularly the case for less liquid stocks, which could result in higher OTR for them, implying that HFT activity is higher for less liquid stocks than for blue chips, which is not in line with existing empirical evidence. Finally, the OTR measure does not take into account the speed at which orders are sent. Hence a firm using an algorithm that updates orders every 10 minutes could have a high OTR even though it is not implementing any HFT strategy.

In other words, OTR is rather a measure of message traffic than a measure of HFT. It is a useful metric to assess potential risks linked to trading system overload rather than a method to identify firms carrying out HFT activity. Looking at our data confirms these considerations. HFT firms exhibit significant heterogeneity. The median unweighted order-to-trade ratio is around 18, while the first quartile is around 3 and the third quartile close to 64. This indicates that HFTs are not a homogeneous category, probably due to the different strategies implemented. For investment banks and other traders, however, order-to-trade ratios centre more around the median, as illustrated in C.5.

Dispersion of order-to-trade ratios   C.5

Note: Distribution of order-to-trade ratios by traders by quartile and 10% and 90% percentiles . Source: ESMA.

4) Identification of strategies

More recently, a few papers have looked at the strategies implemented by HFT firms. Using data from NASDAQ-OMX, Hagströmer and Nordén (2013) are able to identify market making HFT and opportunistic HFT (arbitrage and momentum strategies).

This type of approach is useful to identify the extent to which certain HFT business models contribute to market activity, but may be less suited to identify the overall level of HFT activity in equity markets.

HFT identification methods used in this report

As discussed above, a precise identification of HFT activity is difficult to achieve; from an analytical perspective no single method will exactly capture the extent of HFT activity.

Therefore we will present estimates based on a direct HFT identification approach, using a HFT flag, and an indirect identification approach, based on the lifetime of orders.

For the HFT flag approach a list of firms that engage in HFT has been established with reference to the market participants' primary business based on the information available on their websites, on business newspaper articles and on industry events. In certain cases the flagging of firms was also discussed with supervisors. 20 groups (out of a total of 394) were classified as HFTs in this way.

---

[7] More precisely, orders were first de-trended and HOT were defined as traders at the right tail of the distribution, using a 1.25 standard deviation cut-off point, resulting in an order-to-trade ratio of 11.2 for the HOT group (See Investment Industry Regulatory Organization of Canada, 2012).

[8] See Brogaard et al. (2014).

Our identification rule for HFT activity according to the lifetime of orders approach is as follows: if the 10% quickest order modifications and cancellations of a given firm in any particular stock are faster than 100ms, then the trading activity of the firm in that particular stock is considered HFT activity.

There is no rule which threshold would characterise HFT activity in a precise manner. We have therefore carried out robustness checks and will provide an overview of levels of HFT activity under a lifetime of orders approach for a range of time thresholds.

Thus, our approach is to present a range of estimates for HFT activity based on a HFT flag approach and an order lifetime approach. The results based on the HFT flag provide a lower bound for HFT activity, as they do not capture HFT activity by investment banks. The results based on the lifetime of orders are likely to be an upper bound for HFT activity. As noted above, there may be some degree of overestimation of HFT activity under this approach.

Description of dataset

Sample of stocks

A sample of 100 stocks traded in Belgium (BE), Germany (DE), Spain (ES), France (FR), Ireland (IE), Italy (IT), the Netherlands (NL), Portugal (PT) and the United Kingdom (UK) has been chosen. A stratified sampling approach has been used. For each country, stocks have been split by quartiles according to their market value, value traded and fragmentation using September 2012 data[9]. As in Degryse et al. (2011), fragmentation ($Frag_{i,t}$ of stock i is on day t) is defined as:

$$Frag_{i,t} = 1 - H_{i,t}$$

where $H_{i,t}$ is the Herfindahl-Hirschman index[10].

A random draw was performed to select stocks for each quartile. In order to account for the relative size of the markets, greater weight has been put on larger countries. At the same time, each country in the sample has at least five different stocks.

Sample of stocks by country			C.6
Country	Number of stocks	Country	Number of stocks
BE	6	IT	11
DE	16	NL	13
ES	12	PT	5
FR	16	UK	16
IE	5	All sample	100

Note: Number of stocks in the sample.
Source: ESMA.

The sample includes stocks with very different features. During the observation period (May 2013), average value traded ranged from less than EUR 0.1mn to EUR 611mn. In terms of market capitalization, values ranged from EUR 18mn to EUR 122bn during the observation period (average at EUR 8.7bn and median at EUR 2.9bn). The degree of fragmentation is also very different amongst stocks.

Sample stocks statistics									C.7
Country	Value traded (EUR mn)			Market Cap (EUR bn)			Fragmentation Index		
	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
All sample	33.7	611.3	<0.1	8.7	122	<0.1	0.4	0.7	0
BE	45.7	357.1	0.3	24.3	122	0.8	0.4	0.7	<0.1
DE	37.1	611.3	<0.1	8.2	73	<0.1	0.3	0.7	0
ES	42.8	526	2.6	9.6	41.8	0.7	0.3	0.5	<0.1
FR	34.8	497.2	<0.1	7.5	58	0.1	0.4	0.7	0
IE	5.3	184.7	<0.1	3.6	8.1	<0.1	0.2	0.6	0
IT	33.1	300.7	<0.1	6.5	28.2	0.3	0.2	0.6	0
NL	37.3	350.5	0.3	7.7	51	0.4	0.4	0.7	0.1
PT	17.2	143.1	<0.1	5.3	11.4	2	0.3	0.6	0
UK	29.2	290.2	0.1	8.5	71.2	0.4	0.5	0.7	0.1

Note: Monthly average, minimum and maximum for May 2013. For the fragmentation index a value of 0 indicates no fragmentation (all trading is on one venue), whereas higher values indicate that trading is fragmented across several trading venues.
Source: ESMA.

The data collected covers 12 trading venues: NYSE Euronext Amsterdam (XAMS), Brussels (XBRU), Lisbon (XLIS) and Paris (XPAR), Deutsche Börse (XETR), Borsa Italiana (MTAA), London Stock Exchange (XLON), Irish Stock Exchange (XDUB) and the Spanish Stock Exchange (XMCE), BATS Europe, Chi-X Europe and Turquoise.

The trading venues broadly fall into two categories:

- Incumbent exchanges, where trading on stocks with primary listings on that exchange was concentrated prior to MiFID.
- The market entrants BATS Europe, Chi-X Europe[11] and Turquoise.

---

[9] See Annex 4 for further details on the sampling procedure.

[10] The Herfindahl-Hirschman index is computed on the basis of the sum of squared market shares (value traded) per trading venue. A value of 1 indicates no fragmentation (all trading is on one venue), whereas lower values indicate that trading is fragmented across several trading venues. Consequently for the fragmentation index a value of 0 indicates no fragmentation (all trading is on one venue), whereas higher values indicate that trading is fragmented across several trading venues.

[11] BATS Europe and Chi-X Europe merged in 2011 to form BATS Chi-X Europe. They continue to operate separate trading platforms.

Data collection and timespan

Data was collected by ESMA through National Competent Authorities for the month of May 2013. The dataset covers all messages and trades executed on the aforementioned trading venues as well as some additional information for market members, such as the use of colocation, market making and provision of Direct Market Access. The dataset includes around 10.5 million trades and 456 million messages. Message types include new, modified and cancelled orders (see Annex 5 for a full list of message types).

Market participants

For each trading venue, the list of all market members active during May 2013 was requested by ESMA and the National Competent Authorities. The information requested included internal member ID, name of the member and Bank Identification Code (BIC). In the dataset used in this report all market participants have been anonymised[12].

The identification of firms is based on a stratified approach (Box 2):

i)   for each market participant a Unique ID has been created for each venue where he has membership;
ii) if a participant has several accounts on the same venue, each account will have a separate ID but the same Account ID;
iii) if a market participant is a member of several venues, all these accounts will have the same Group ID[13]; and
iv) a Master ID has been created to include all market members that are linked to the same entity.

For the HFT flag approach each market participant is flagged as HFT, investment bank or other. A list of firms that engage in HFT has been established with reference to the market participants' primary business based on information available on their websites, in business newspaper articles and industry events. In certain cases the flagging of firms was also discussed with supervisors; 20 groups (out of a total of 394) were classified as HFTs in this way.

Table C.8 shows the number of active market participants that are flagged as HFT in our sample.

Identification of HFT firms							C.8
Approach			Participants			Groups	
	Indicator	HFT	Non-HFT	Total	HFT	Non-HFT	Total
Direct	List of firms	181	1030	1211	20	374	394
Information items							
Number of investment banks			319			57	
Number of other market participants			711			317	

Note: Each market member is identified as a unique participant ID; if several market members have the same name they have the same Group ID.
Source: ESMA.

On average, HFT groups have 9.1 different unique IDs, indicating that their trading is spread across multiple venues. Investment banks have 5.6 different IDs and other firms 2.2. Those features are in line with the assumption that HFTs are more likely to perform arbitrage across venues than other types of market participants. However, as shown in Table C.9, there is substantial variation within each group, especially for investment banks and other members.

Number of IDs by types				C.9
Type	Average	Median	Max	Min
HFT	9.1	10	13	1
IB	5.6	3	23	1
Other	2.2	1	13	1
All	3.1	1	23	1

Note: Number of IDs within HFT groups, IB groups or other groups across all trading venues in sample.
Source: ESMA.

---

[12] The data have been subject to two anonymisation procedures. First, the providers of the data (either trading venues or National Competent Authorites) have transmitted data where traders are identified using trading venue specific anonymous codes. Datasets have then been merged using correspondence tables ensuring the consistency of the trading firms identification across venues. Finally, new anonymised IDs have been created for each trading firm by ESMA.

[13] Groups were established based on the name of the market members, see Annex 3 for further details.

HFT activity on European equity markets

We assess the extent of HFT activity on European equity markets by looking at the share of HFT activity in terms of the value traded of shares, the number of trades and the number of orders.

This section first describes the overall estimates for HFT activity in our sample and provides robustness checks regarding the lifetime of orders approach. Results for the direct approach (lower bound) and the indirect lifetime of orders approach (upper bound) are provided separately.[14] The remainder provides more detailed analysis of our results, focussing on differences in HFT activity between types of market participants, the use of colocation, patterns in HFT activity and any relationship between HFT activity and the underlying features of the stocks in our sample.

Overall results for HFT activity

Overall, HFT firms account for 24% of value traded in our sample, based on the HFT flag approach. Based on the lifetime of orders approach, HFT activity accounts for 43% of value traded.

For the number of trades, the corresponding numbers are between 30% for the HFT flag approach and 49% for the lifetime of orders approach; for the number of orders they are between 58% and 76%, respectively.

The difference is mainly explained by the activity of investment banks. They account for around 61% of total value traded, of which roughly one third (22% of total value traded) is identified as HFT activity in a lifetime of orders approach (see C.15).

Identification approach	Value traded	Number of trades	Number of orders
HFT flag	24	30	58
Lifetime of orders	43	49	76

HFT activity - overall results   C.10

Note: Figures are weighted by value of trades (value traded), number of trades, and number of orders in %.
Source: ESMA.

Across all venues, the share of HFTs by value traded was smaller than the share by number trades, which in turn was lower than the HFT share by number of orders. This indicates firstly that the size of HFT trades is smaller than the size of non-HFT trades. Moreover, it indicates that the order-to-trade ratio of HFTs is on average higher than order-to-trade ratio of non-HFTs.

HFT activity varies significantly between trading venues. In terms of value traded, HFT activity ranges from 8% to 40% (average 24%) for the HFT flag approach and from 19% to 63% (average 43%) for the lifetime of orders approach. For number of trades, HFT activity ranges between 9% and

44% (average 30%) for the HFT flag approach and between 18% and 65% (average 49%) for the lifetime of orders approach. For number of orders the range for HFT activity is between 31% and 76% (average 58%) for the HFT flag approach and between 34% and 87% (average 76%) for the lifetime of orders approach.[15]

Overview of HFT activity - HFT flag and lifetime of orders   C.11

Trading venue	Value traded		Number of trades		Number of orders	
	HFT flag	Lifetime of orders	HFT flag	Lifetime of orders	HFT flag	Lifetime of orders
All venues	24	43	30	49	58	76
BATE	40	60	44	63	76	85
CHIX	40	56	40	58	59	80
MTAA	25	20	26	18	51	34
TRQX	34	63	35	65	73	84
XAMS	24	48	28	54	53	77
XBRU	18	48	23	50	38	64
XDUB	8	19	9	28	43	87
XETR	21	35	24	35	33	63
XLIS	11	40	17	45	31	65
XLON	21	32	26	35	44	56
XMCE*	0	32	0	29	0	46
XPAR	21	45	30	51	50	70

Note: Figures are weighted by value of trades (value traded), number of trades and number of orders, in %. For trades on UK stocks, value traded has been converted to EUR using end-of-day exchange rates.

BATE=BATS, CHIX=Chi-X MTAA= Borsa Italiana, TRQX=Turquoise, XAMS=NYSE Euronext Amsterdam, XBRU=NYSE Euronext Brussels, XDUB=Irish Stock Exchange, XETR=Deutsche Boerse AG, XLIS=NYSE Euronext Lisbon, XLON=London Stock Exchange, XMCE=Mercado Continuo Español, XPAR=NYSE Euronext Paris.

*No HFT firms were direct members of XMCE during the observation period. Therefore no HFT activity is reported for XMCE under the HFT flag approach.

Source: ESMA.

Lifetime of orders approach – robustness of results regarding the time threshold

As described earlier, there is no general rule which threshold characterises HFT activity in a precise manner. We therefore carry out robustness checks and analyse levels of HFT activity under a lifetime of orders approach for a range of time thresholds from 5ms to 500ms. As expected, results for HFT activity increase with longer time thresholds for the lifetime of orders. There are initially

---

[14] Annex 2 provides results for a few other identification methods; message traffic as well as lifetime of orders approach calculated at firm level and based on the relative lifetime of orders at group level compared to lifetime of orders of the trading venue.

[15] On 15 May 2013 the German HFT Act came into force. Comparing HFT activity during the time before and after the Act came into force does not show any clear patterns. In terms of value traded estimated HFT activity under the HFT flag approach on Deutsche Boerse was lower after the introduction of the act (19% compared to 23% before). Across all venues there was also a decrease in HFT activity, albeit lower (23.4% compared to 23.8%).
However, under the lifetime of orders approach the picture is different. Here, HFT activity on Deutsche Boerse is higher after the introduction of the act (36% compared to 34% before) whereas HFT activity across venues was stable at 43%.
The picture regarding number of trades is comparable to the observations for value traded. For the number of orders, both the HFT flag and the lifetime of orders approach show slightly increased HFT activity on Deutsche Boerse after the introduction of the HFT act.

strong increases until a lifetime of orders of about 40ms, after which the curve becomes flatter.

For the number of orders, HFT activity is between 84% for a threshold of 500ms and 72% for a threshold of 50ms. For number of trades, HFT activity varies between 64% for 500ms and 44% for 50ms, for value traded it varies between 60% and 37%, respectively. This indicates that there is some variation in results for HFT activity, depending on the choice of time threshold for the lifetime of orders. One should also note that for all thresholds of 20ms and above estimated HFT activity under the lifetime of orders approach is higher than under the HFT flag approach.

Lifetime of orders approach - robustness of results   C.12

Note: Share of HFT activity, in %, using different thresholds, in ms, for the lifetime of orders.
Source: ESMA
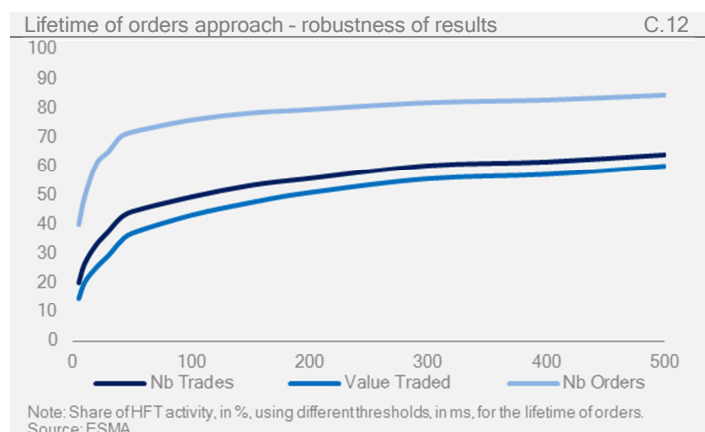
Activity by different type of market participants

The HFT flag approach allows us to separate the overall levels of trading activity by HFT firms, investment banks and other market participants.

In terms of value traded HFT firms account for 24% of overall trading activity, investment banks for 61% and other firms for 15%. In terms of number of trades the respective shares are 30% for HFT firms, 59% for investment banks and 12% for other firms. Regarding the number of orders HFT firms account for a higher share, 58% of the overall number of orders, whereas investment banks account for 39% and other firms for 3% of the overall number of orders.

Trading activity by type of market participant   C.13

Note: Activity by trader type in %.
Source: ESMA
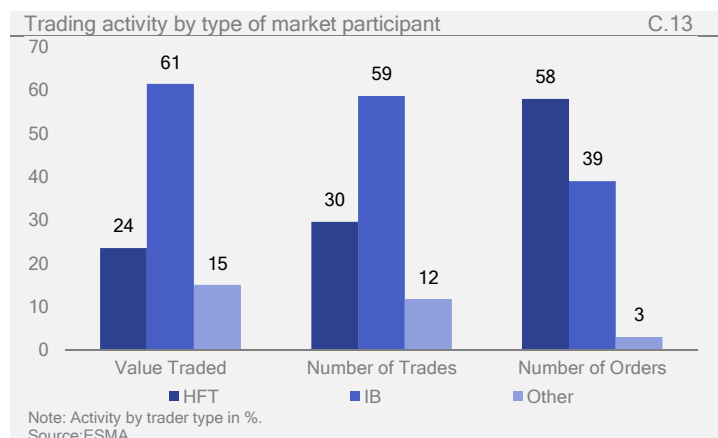
Using both the HFT flag approach and a lifetime of orders approach to identify HFT activity allows us to estimate the overall level of HFT activity in our sample and at the same

time explain which type of market participant acts as HFT. To our knowledge, previous studies have looked at one of these aspects at a time, but have not analysed these two dimensions of HFT activity together.

HFT activity under the lifetime of orders approach is concentrated in HFT firms and investment banks. Investment banks show higher HFT activity in terms of value traded than firms classified as HFT firms. There is little HFT activity by other firms under the lifetime of orders approach. 43% of value traded results from HFT activity under the lifetime of orders approach, split between 19% provided by HFT firms, 22% by investment banks and 2% by other market participants. For the number of trades 49% result from HFT activity: 25% by HFT firms, 23% by investment banks and 2% by other market participants. For the number of orders 76% result from HFT activity; 55% by HFT firms, 19% by investment banks and 1% by other market participants.

HFT activity -results by market participants   C.14

	HFT flag	Lifetime of orders		
		Lifetime of orders: total	thereof HFT firms	thereof investment banks	thereof other firms
Value traded	24	43	19	22	2
Number of trades	30	49	25	23	2
Number of orders	58	76	55	19	1

Note: Figures are weighted by value of trades (value traded), number of trades, and number of orders in %.
Source: ESMA.

Looking at the activity of investment banks in more detail shows that the level of both their overall and HFT trading activity varies widely between trading venues.

Under the HFT flag approach investment banks account for 61% of value traded (from 20% to 75% depending on the trading venue), 59% of number of trades (from 29% to 69%) and 39% of the number of orders (from 24% to 70%).

Under the lifetime of orders approach, HFT activity by investment banks accounts for 22% of overall value traded (from 10% to 34% depending on the trading venue), 23% of overall number of trades (from 7% to 34%) and 19% of the overall number of orders (from 9% to 35%).

Investment banks - total and HFT activity					C.15
Trading venue	Value traded		Number of trades		Number of orders	

Trading venue	Total activity	HFT activity	Total activity	HFT activity	Total activity	HFT activity
All venues	61	22	59	23	39	19
BATE	55	25	52	25	24	12
CHIX	55	22	55	23	39	23
MTAA	42	10	39	7	39	9
TRQX	64	34	62	34	26	13
XAMS	65	25	60	26	45	24
XBRU	75	30	69	26	61	26
XDUB	20	10	29	15	24	22
XETR	67	20	62	19	54	24
XLIS	58	29	58	28	65	35
XLON	73	16	69	16	50	14
XMCE	59	22	62	20	70	27
XPAR	71	24	62	20	50	20

Note: Figures are weighted by value of trades (value traded), number of trades and number of orders, in %.
BATE=BATS, CHIX=Chi-X MTAA= Borsa Italiana, TRQX=Turquoise, XAMS=NYSE Euronext Amsterdam, XBRU=NYSE Euronext Brussels, XDUB=Irish Stock Exchange, XETR=Deutsche Boerse AG, XLIS=NYSE Euronext Lisbon, XLON=London Stock Exchange, XMCE=Mercado Continuo Español, XPAR=NYSE Euronext Paris,
Source: ESMA.

The lifetime of orders approach also allows us to separate the trading activity of different market participants into HFT activity and non-HFT activity (see C.16).

HFT firms account for 24% of value traded, with 19% classified as HFT activity and 5% as non-HFT activity under the lifetime of orders approach. For the number of trades HFT firms account for 30% of trading activity, with 25% classified as HFT activity and 5% as non-HFT activity. For the number of orders they account for 58% of overall order activity, 55% are classified as HFT activity and 3% as non-HFT activity.

These results suggest that market participants classified as HFT firms act as HFTs in the majority of stocks they are active in, but their trading activity is slower in other stocks. This is corroborated by the observation that the proportion of orders by HFT firms classified as HFT activity is higher than the proportion of HFT activity for value traded and number of trades. Thus their order-to-trade ratio appears to be higher for their HFT activity compared to their non-HFT activity.

Investment banks account for 61% of overall value traded. 22% are classified as HFT activity and 39% as non-HFT activity. For number of trades, investment banks account for 59% of overall activity, with 23% classified as HFT activity and 36% as non-HFT activity. For number of orders, they account for 39% of overall orders in our sample, with 19% classified as HFT activity and 20% as non-HFT activity.

As for HFT firms, the proportion of investment banks' HFT activity is higher for number of orders than for value traded and number of trades. Therefore also for investment banks the order-to-trade ratio appears to be higher for their HFT activity compared to their non-HFT activity.

Other firms account for 15% of overall value traded, 2% of which is classified as HFT activity and 13% as non-HFT activity. In terms of number of trades other firms account for 12% of overall activity, 2% is classified as HFT activity and 10% as non-HFT activity. For number of orders, they account for 3% of overall orders in the market, 1% are classified as HFT activity and 2% as non-HFT activities.

HFT and non-HFT trading activity by HFT firms, investment banks and other firms				C.16
		Value traded	Number of trades	Number of orders
HFT firms		24	30	58
Thereof	HFT activity	19	25	55
	Non-HFT activity	5	5	3
Investment banks		61	59	39
Thereof	HFT activity	22	23	19
	Non HFT activity	39	36	20
Other firms		15	12	3
Thereof	HFT activity	2	2	1
	Non-HFT activity	13	10	2

Note: Figures are weighted by value of trades (value traded), number of trades and number of orders, in %.
Source: ESMA.

Colocation

We also looked at the use of colocation services as a proxy for HFT activity. In our sample, most of the HFT activity is linked to market participants using colocation services. Colocated market participants account for around 75% of value traded and number of trades as well as 92% of the number of orders.[16]

As shown in C.17, firms identified as HFT using the HFT flag approach account for 35% of colocation users, investment banks for 44% and other market participants for 21%.

Use of colocation services   C.17

Others 21%
HFT 35%
Investment Banks 44%

Note: Share of colocation users by type
Source: ESMA
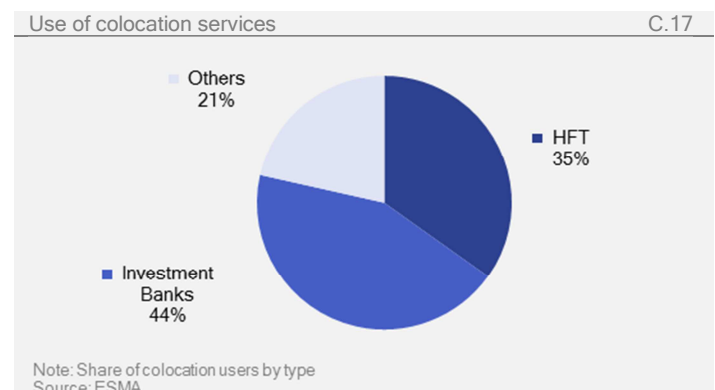
[16] For three trading venues the information was not provided as colocation services were outsourced to third parties.

Within the different categories of market participants, 80% of HFT groups use colocation in at least one trading venue, against 37% for investment banks and 9% for other firms.

These results show that the use of colocation services is not a good proxy for HFT activity. It would capture both HFT and non-HFT activity of investment banks and other market participants and thus overestimate HFT activity.

Patterns in HFT activity

In this section we look at patterns in HFT activity across our observation period of May 2013. This includes patterns we observe in intraday activity in general and specifically for HFT activity during auctions and continuous trading.

HFT activity in terms of value traded was relatively stable during the observation period of May 2013.[17] Under the HFT flag approach, median daily HFT activity for the sampled stocks ranges from 21% to 30%, as shown in C.18. The lowest values were observed at the beginning and at the end of the month.[18]. Under the lifetime of orders approach, median daily HFT activity ranges from 31% to 52% (see C.19).

Daily HFT activity - HFT flag   C.18

Note: Mediam, first and third quartiles for HFT activity, in %, based on value traded. For readibility, 28 stocks with monthly average HFT activity higher than 20% have been used. Source: ESMA.

Daily HFT activity - lifetime of orders   C.19

Note: Median, first and third quartiles for HFT activity, in %, based on value traded in the same 28 stocks as used in chart C.18. Source: ESMA.

HFT activity on an intraday basis can be quite different depending on the stock and day. In C.20, for example, aggregated HFT activity for a stock is relatively stable during the trading day at around 45% of value traded for the flag approach and around 50% of value traded for the lifetime of orders approach.

In contrast the HFT activity in another stock (C.21) is very volatile both under the HFT flag and the lifetime of orders approach. HFT activity is increasing from 25% to 70% around 12:30 under the HFT flag approach.

A common feature for both stocks is the drop in HFT activity at the end of the trading day. This indicates that HFT firms tend to avoid auctions. The drop in HFT activity at the end of the trading day is less marked under the lifetime of orders approach.

Stable intraday HFT activity   C.20

Note: Number of trades by HFT and non-HFT firms in one liquid stock for one day, consolidated across venues and aggregated every 10 minutes. HFT share in number of trades, in %. Source: ESMA.

Volatile intraday HFT activity   C.21

Note: Number of trades by HFT and non-HFT firms in one liquid stock for one day, consolidated across venues and aggregated every 10 minutes. HFT share in number of trades, in %. Source: ESMA.

Looking at this observation in more detail, we find that HFT activity overall accounts for 24% of value traded under the HFT flag approach, whereas during auctions its share amounts only to 3%. In terms of number of trades the picture is similar, the respective shares are 30% for the overall number of trades and 3% during autions. This could be explained by the inventory management of HFTs which tend to manage their position intraday. As a result they may aim to have a neutral position before the closing auction.

Trading activity during auctions accounts for around 17% of value traded and around 2.5% of the number of trades. As a result, excluding auctions, HFT activity under the HFT flag approach would be around 28% for value traded and 30% for the number of trades.

---

[17] The German HFT Act came into force during our observation period on 15 May 2013. As mentioned above we have not observed any clear patterns regarding changes in HFT activity before and after the introduction of the German HFT act in our sample.

[18] The lower HFT activity on 1st May can be explained by the fact that most trading venues were closed due to a bank holiday, with the exception of XLON and especially XDUB, where there is very low HFT activity.

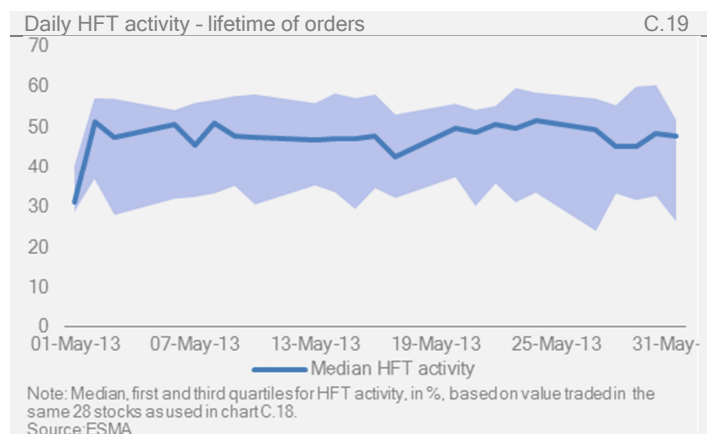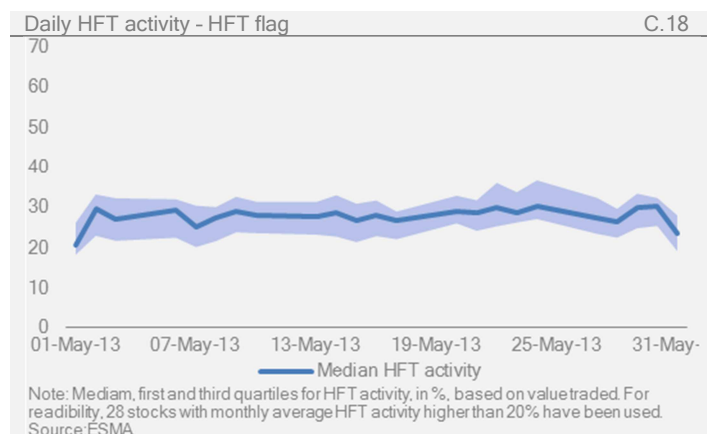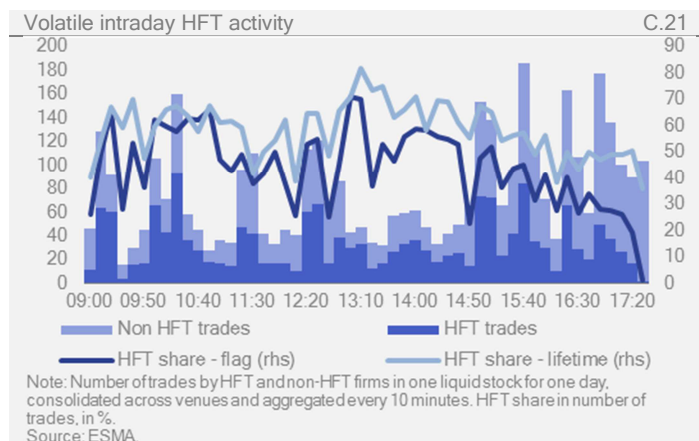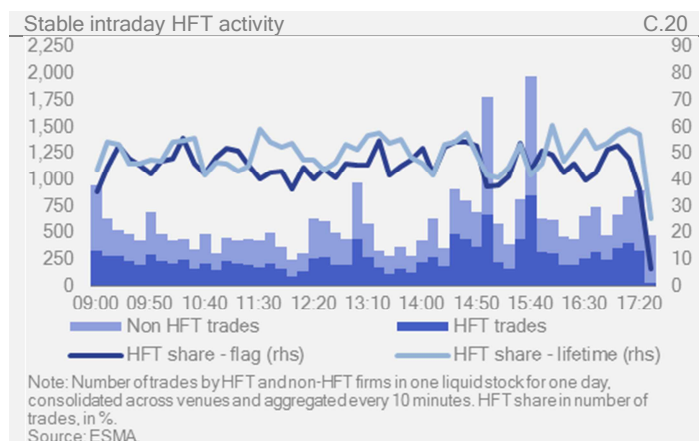| HFT activity - flag and lifetime data				C.22
Trading venue	Value traded		Number of trades	
	HFT flag	Lifetime of orders	HFT flag	Lifetime of orders
All venues	24	43	30	49
All venues auction only	3	32	3	29
All venues excluding auctions	28	45	30	50

Note: Figures are weighted by value of trades (value traded) and number of trades, in %. For trades on UK stocks, value traded has been converted to EUR using end-of-day exchange rates.
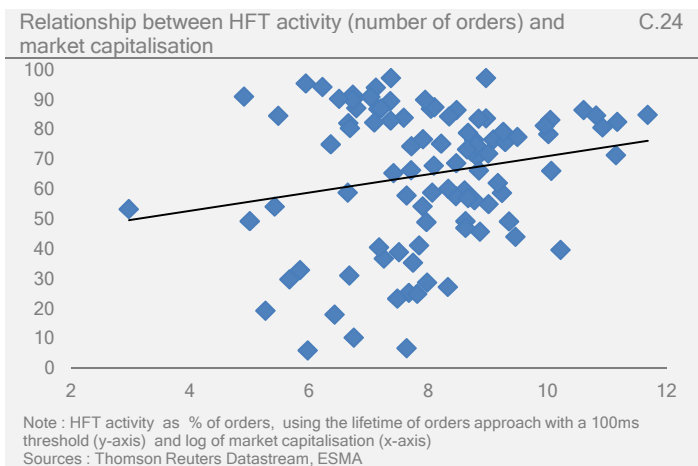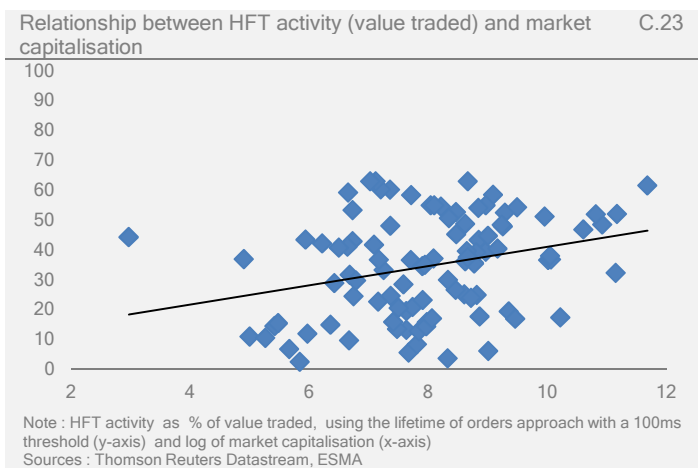
Source: ESMA.

Using the lifetime of orders approach, HFT activity is still lower during auctions, but the result is less marked. By value traded HFT activity is 32% compared to 43% of overall HFT activity. By number of trades the numbers are 29% and 49% respectively.

This can be explained by one of the limitations of the lifetime of orders approach. The lifetime of orders approach also captures HFT activity by investment banks. However, we might overestimate their HFT activity in the following case. Where the trading activity of an investment bank in stock X is classified as HFT activity according to our rule, not all of the trading activity may in fact be HFT activity (we could capture both the activity of a HFT desk and the slower agent-based trading activity). The share of HFT activity during auctions being closer to the overall share of HFT activity under the lifetime of orders approach is an indication that this effect is indeed present to some extent.

HFT activity and underlying stocks' features

Empirical work on HFT indicates that HFTs tend to trade stocks with high market value ('blue chips')[19], due to higher liquidity.

This can also be seen in our dataset. C.23 and C.24 present the simple relationship between HFT activity in terms of value traded and market value of stocks in our sample as well as the relationship between number of HFT orders and market value. We are aware that other factors such as volatility, liquidity and fragmentation also have an impact on this relationship.

Relationship between HFT activity (value traded) and market capitalisation   C.23

Note : HFT activity as % of value traded, using the lifetime of orders approach with a 100ms threshold (y-axis) and log of market capitalisation (x-axis)
Sources : Thomson Reuters Datastream, ESMA

Relationship between HFT activity (number of orders) and market capitalisation   C.24

Note : HFT activity as % of orders, using the lifetime of orders approach with a 100ms threshold (y-axis) and log of market capitalisation (x-axis)
Sources : Thomson Reuters Datastream, ESMA

Conclusion

This report describes the results of the first part of ESMA's research on HFT. It complements the existing HFT literature in two ways.

Firstly, it provides estimates on HFT activity across EU equity markets. Most of the HFT studies published so far focus either on the US or on a single country within Europe. We provide estimations for HFT activity based on the primary business of firms (direct approach) as well as based on the lifetime of orders (indirect approach). The results based on the primary business of firms provide a lower bound for HFT activity, as they do not capture HFT activity by investment banks, whereas the results based on the lifetime of orders are likely to be an upper bound for HFT activity.

Secondly, using both a direct HFT identification approach (primary business of firms) and an indirect identification approach (lifetime of orders) allows us to estimate the overall level of HFT activity in our sample and at the same time explain which type of market participant acts as HFT.

In our sample, we observe that HFT activity accounts for 24% of value traded for the HFT flag approach and 43% for the lifetime of orders approach. For the number of trades the corresponding numbers for HFT activity are 30% and 49%, and for the number of orders 58% and 76%.

The difference in the results is mainly explained by significant HFT activity of investment banks, which is captured under a lifetime of orders approach, but not

---

[19] See for example Brogaard, J., T. Hendershott, and R. Riordan (2014).

under a HFT flag approach. We also find that the level of HFT activity varies between trading venues.

Regarding the characteristics of market participants we find that HFT firms are members of more trading platforms than other types of market participants, which amongst other reasons may indicate that they are more likely to perform cross-venue arbitrage.

In more general terms, our results show that depending on the identification approach chosen, the estimated level of HFT activity varies significantly. This remains an important issue for the analysis of HFT activity and its impacts.

This report describes the results of the first part of the ESMA research on HFT. Further research is needed

— regarding the drivers of HFT activity,

— to assess the actual contribution of HFT to liquidity, and

— to analyse potential risks and benefits linked to HFT activity.

References

Aitken, M., D. Cumming, and F. Zhan, 2014. Trade Size, High Frequency Trading, and Co-Location around the World. European Journal of Finance, forthcoming.

Alampieski, K., and A. Lepone, 2011. High Frequency Trading Firms, Order Book Participation and Liquidity Supply during Periods of Heightened Adverse Selection Risk: Evidence from LSE, BATS, and Chi-X. Mimeo.

———, 2012. High Frequency Trading in UK Equity Markets: Evidence Surrounding the US Market Open. Mimeo.

Aldridge, I., 2010. High-frequency trading: a practical guide to algorithmic strategies and trading system, Wiley Trading Series.

Australian Securities & Investments Commission, 2013. Dark Liquidity and High-Frequency Trading. Report 331.

Baron, M., J. Brogaard, and A. Kirilenko, 2012. The Trading Profits of High-Frequency Traders.

Benos, E., and S. Sagade, 2012. High-Frequency trading behavior and its impact on market quality: evidence from the UK equity market. Working Paper No. 469.

Bershova, N., and D. Rakhlin, 2013. High-Frequency Trading and Long-Term Investors: A View from the Buy-Side. Journal of Investment Strategies, 2. 25–69.

Bessembinder, H., and K. Venkataraman, 2010. Bid-Ask Spreads: Measuring Trade Execution Costs in Financial Markets, in Encyclopedia of Quantitative Finance, edited by R. Cont, John Wiley and Sons.

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

Boehmer, E., K. Fong, and J. Wu, 2012. International Evidence of Algorithmic Trading.

Boehmer, E., and R.L. Shankar, 2014. Low Latency Trading and the Comovement of Order Flow, Prices and Market Conditions.

Breckenfelder, H.J., 2013. Competition between High-Frequency Traders and Market Quality.

Brogaard, J., 2010. High Frequency Trading and Its Impact on Market Quality.

———, 2011. The Activity of High Frequency Traders.

Brogaard, J., C. Garriot, and A. Pomeranets, 2014. High-Frequency Trading Competition. Bank of Canada Working Paper, no, 2014-19.

Brogaard, J., B. Hagstromer, L. Norden, and R. Riordan, 2013. Trading Fast and Slow: Colocation and Market Quality.

Brogaard, J., T. Hendershott, S. Hunt, and C. Ysusi, 2014. High-frequency trading and the execution costs of institutional investors. Financial Review, 49(2): 345-369.

Brogaard, J., T. Hendershott, and R. Riordan, 2014. High Frequency Trading and Price Discovery. Review of Financial Studies.

Carrion, A., 2013. Very Fast Money: High-Frequency Trading on the NASDAQ. Journal of Financial Markets 16 (4): 680–711.

Clark-Joseph, A., 2013. Exploratory Trading, mimeo.

Commodity Futures Trading Commission, 2012. Presentation by the CFTC Technology Advisory Committee Subcommittee on Automated and High Frequency Trading, 2 June 2012.

Credit Suisse, 2012. High Frequency Trading-Measurement, Detection and Response, Market Commentary, 6 December 2012.

Degryse, H., F. de Jong, and V. van Kervel, 2011. The impact of dark trading and visible fragmentation on market quality, CEPR discussion paper 8630, November.

ESMA, 2011. Guidelines on systems and controls in an automated trading environment for trading platforms, investment firms, and competent authorities.

Gai, J., C. Yao, and M. Ye, 2013. The Externalities of High-Frequency Trading.

Gao, C., and B. Mizrach, 2011. High Frequency Trading in the Equity Markets during Large Scale Asset Purchases.

———, 2013. Market Quality Breakdowns in Equities.

Gerig, A., 2012. High-Frequency Trading Synchronizes Prices in Financial Markets.

Gomber, P., B. Arndt, M. Lutat, and T. Uhle, 2011. High-Frequency Trading, Report commissioned by Deutsche Börse Group.

Gomber, P., and M. Gsell, 2009. Algorithmic Trading Engines versus Human Traders - Do they behave differently in securities markets?, Center for Financial Studies working paper No. 2009/10, April.

Gresse, C., 2010. Multi-Market Trading and Market Quality, mimeo.

Grillet-Aubert, L., 2010. Equity trading: a review of the economic literature for the use of securities market regulators, AMF working paper No. 9, June.

Haasbrouck, J., and G. Saar, 2013. Low-Latency Trading, Johnson School Research Paper Series No. 35-2010.

Hagströmer, B., and L. Nordén, 2013. The diversity of high frequency traders, Journal of Financial Markets, Vol. 16(4). November.

Hagströmer, B., L. Nordén, and D. Zhang, 2013. How Aggressive Are High-Frequency Traders?

Hasbrouck, J., and G. Saar, 2013. Low-Latency Trading.

Hasbrouck, J., and G. Saar, 2009. Technology and Liquidity Provision: The Blurring of Traditional Definitions. Journal of Financial Markets 12 (2): 143–72.

Hendershott, T., C. Jones, and A.J. Menkveld, 2011. Does Algorithmic Trading Improve Liquidity?, Journal of Finance, Vol. 66(1). Febuary.

Hendershott, T., and R. Riordan, 2011. Algorithmic Trading and Information.

———, 2013. Algorithmic Trading and the Market for Liquidity, Journal of Financial and Quantitative Analysis, forthcoming.

Hirschey, N.H., 2013. Do High-Frequency Traders Anticipate Buying and Selling Pressure?.

Huh, Y., 2014. Machines vs. Machines: High Frequency Trading and Hard Information.

Investment Industry Regulatory Organization of Canada, 2012. The HOT Study - Phases I and II of IIROC's Study of High Frequency Trading Activity on Canadian Equity Marketplaces.

Jarnecic, E., and M. Snape, 2014. The provision of liquidity by high frequency participants. Financial Review 49(2): 371-394.

Johnson, H., B.F. Van Ness, and R.A. Van Ness, 2014. An Analysis of All, Pure, and Circumstantial Odd-Lot Transactions.

Jones, C., 2013. What Do We Know About High-Frequency Trading?, Columbia Business School Research Paper No. 13-11.

Jovanovic, B., and A.J. Menkveld, 2010. Middlemen in Limit-Order Markets, mimeo.

Kang, J., and J. Shin, 2012. The Role of High Frequency Traders in Electronic Limit Order Markets.

Kirilenko, A., A. Kyle, M. Samadi, and T. Tuzun, 2010. The flash crash: The impact of high frequency trading on an electronic market, Working Paper.

Kirilenko, A., and A. Lo, 2013. Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents, Journal of Economic Perspectives, Vol. 27, 2.

Malinova, K., A. Park, and R. Riordan, 2013. Do retail traders suffer from high frequency traders?, working paper

Menkveld, A.J., 2013. High Frequency Trading and the New-Market Makers. Journal of Financial Markets 16 (4): 712–40.

Menkveld, A.J., and M.A. Zoican, 2014. Need for Speed? Exchange Latency and Liquidity.

O'Hara, M., C. Yao, and M. Ye, 2012. What's Not There: The Odd-Lost Bias in TAQ Data.

Pagano, M., 1989. Trading volume and asset liquidity, Quarterly Journal of Economics, Vol. 104.

Scholtus, M., and D.J.C. Van Dijk, 2012. High-Frequency Technical Trading: The Importance of Speed. Tinbergen Institute Discussion Paper, no. 12-018/4.

Scholtus, M., D.J.C. Van Dijk, and B. Frijns, 2012. Speed, Algorithmic Trading and Market Quality around Macroeconomics News Announcements.

Tong, L., 2013. A Blessing or a Curse? The Impact of High Frequency Trading on Institutional Investors.

U.S. Securities and Exchange Commission, 2010. Part III: Concept Release on Equity Market Structure; Proposed Rule, 17 CFR Part 242, Federal Register, Vol. 75, No. 13, January 21, 2010, Proposed Rules, pages 3594-3614. http://www.sec.gov/rules/concept/2010/34-61358fr.pdf.

U.S. Securities and Exchange Commission, 2014. Equity Market Structure Literature Review. Part II: High Frequency Trading, March 18, 2014. http://www.sec.gov/marketstructure/research/hft_lit_review_march_2014.pdf

Weller, B., 2012. Liquidity and High Frequency Trading, mimeo.

Zhang, S.S., 2010. High-Frequency Trading, Stock Volatility, and Price Discovery.

———, 2013. Need for Speed: An Empirical Analysis of Hard and Soft Information in a High-Frequency World.

Zhang, S.S., and R. Riordan, 2011. Technology and Market Quality: The Case of High Frequency Trading.

Annex 1: Literature review

This literature review serves two purposes. Firstly, it provides an overview of the datasets used in the identification of HFT. Secondly, it reviews HFT identification strategies employed in the literature.

Datasets

The list of available HFT datasets has been growing considerably over the last years. The extent to which the data enable HFT identification is, however, variable. A common limitation of datasets is data frequency, which is too low to observe HFT activity. Another common issue is the unavailability of order-level data, which are necessary to compute order lifetimes or order-to-trade ratios.

This section surveys datasets that were used in empirical studies of HFT activity or identification. The datasets identified have been grouped into four categories, based on the geographic region of the studied trading venues. These categories are (1) EU datasets, (2) US datasets, (3) other countries and (4) cross-country datasets, which contain data from venues in at least three countries in different geographic regions. As evident from C.1, US datasets are most often used, followed by European datasets. Few authors study datasets from other countries – notably Australia and Canada – and cross-country datasets are rare.

Geographic region of datasets in the HFT literature   C.1

Source: ESMA
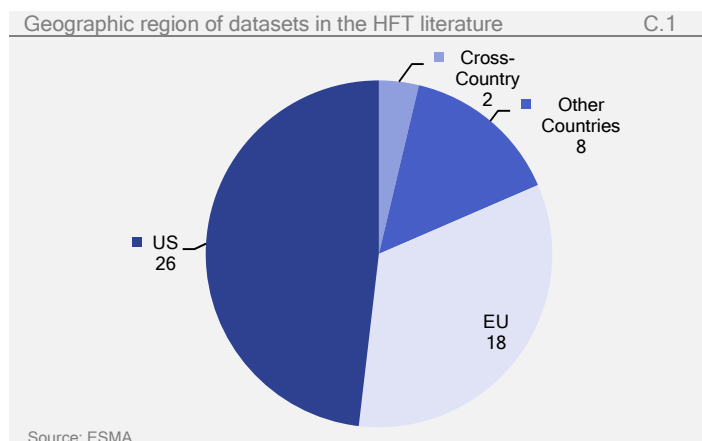
EU Datasets

Alampieski and Lepone (2011; 2012) study a dataset provided by the UK Financial Service Authority (FSA), containing trade and order book data from three venues in the UK: the London Stock Exchange (LSE), Chi-X Europe and BATS Europe (prior to the merger between the two). The dataset contains all constituent stocks in the FTSE 100 for 30 trading days in 2010. The authors focus solely on a subsample of 22 stocks that are cross-listed on American exchanges, however. Trading data are aggregated on firm-level, however, thus not allowing a differentiation between different types of HFTs. Timestamps are to the closest second.

Other studies using UK FSA transaction reporting data are Benos and Sagade (2012) and Brogaard, Hendershott, et al. (2014). Those data contain information on transaction price, size, date and time (reported to the closest second), location as well as the counterparty identity. The latter identifies both buyer and seller in each transaction as well as whether the trade has been executed on behalf of an agent (Benos and Sagade 2012, 3). While the database excludes most transactions by firms not directly regulated by the FSA or FCA, it does "include the trades of some of the largest HFTs". Some of the activity of these HFT firms can be indirectly identified if the HFT firms are reported as counterparties or clients.

Jarnecic & Snape (2014) obtained a dataset by the London Stock Exchange. Their data cover all FTSE100 stocks in three sample periods; April to June 2009, June 2007 and June 2008. The latter two samples are used for robustness checks. Their data include all order book messages, i.e. entries, amendments and cancellations, allowing them to construct the limit order book at any point within the sample horizon.

NASDAQ OMX Nordic offers access to high frequency data-feeds on its Nordic and Baltic venues. Breckenfelder (2013), Hagstromer and Norden (2013), Hagstromer, Norden and Zang (2013) and Brogaard, Hagstromer, et al. (2013) use datasets of the Swedish stock exchange (NASDAQ OMXS). The dataset contains all trading information for trades executed on the exchange, with timestamps to the closest millisecond. Brogaard, Hendershott, et al. (2013) explain that all messages entered in the INET system of OMXS are available via the ITCH feed, including limit order submissions, cancellations, modifications and executions. For each item, data fields are provided identifying time of entry, quantity, limit price, trader identity information, visibility conditions and time in force as well as an order sequence number.

Menkveld and Zoican (2014) use the Thomson Reuters Tick History (TRTH) database to construct a cross-venue sample of NASDAQ OMX exchanges in Copenhagen, Helsinki and Stockholm. Focusing on the 40 stocks included in the OMX Nordic 40 index, TRTH provides trade and quote information with trader identity revealed for both sides of the transaction.

Xetra data from Deutsche Boerse AG are used in AT and HFT studies by Gomber and Gsell (2009), Gomber et al. (2011) and Hendershott and Riordan (2011; 2013). These typically contain all order book events, with timestamps on a 10 microsecond-basis. A detailed description of Xetra data can be found in the appendix of Hendershott and Riordan (2011).

Menkveld (2013) and Jovanovic and Menkveld (2010) obtained access to two very similar datasets, with detailed trading data for both Chi-X and Euronext. Jovanovic and Menkveld (2010) use a sample of all Dutch nonfinancial index stocks on Chi-X and Euronext. Menkveld (2013) uses data of 14 Dutch stocks and 18 Belgian stocks on the same two locations. The datasets contain transaction price, size and an anonymised broker ID for both sides of the transaction. In Menkveld (2013) timestamps are to the second on Euronext and to the millisecond on Chi-X. The data in Jovanovic and Menkveld (2010) are time-stamped to the second.

US Datasets

The most prevalent datasets in the empirical literature are the HFT datasets provided by NASDAQ. At present, there appears to be a total of fourteen studies that use these datasets.[20] A thorough description of the datasets can be found in Brogaard, Hendershott, and Riordan (2014). The data cover a sample of 120 randomly selected stocks listed on NYSE and NASDAQ, where NYSE data were retrieved via the TAQ database (see below). The stocks are categorised into equal tranches based on market capitalization, 40 large-caps, 40 mid-caps and 40 small-cap stocks. The allocation is based on the rank by market capitalization in the Russell 3000 index. Small-caps are stocks around the 2000th by market cap, mid-cap around the 1000th and large-caps are amongst the largest market capitalization stocks in the index. The tranches are equally divided amongst the two venues, with 20 stocks of each tranche listed on NYSE and NASDAQ respectively. Trade data are available for the years 2008 and 2009 and are time-stamped to the millisecond. For each trade, the sample identifies each counterparty as HFT or non-HFT and whether the trade was buyer or seller initiated. In doing so, NASDAQ's HFT dataset is amongst the few datasets that pre-flags HFT firms.

NASDAQ TotalView ITCH subscriptions grant access to daily recordings of direct feed data from NASDAQ exchanges. These so called ITCH files are available not only for US exchanges, but also for non-US NASDAQ trading venues. Data elements include order level data and trade messages, administrative messages and net order imbalance data. The potential timestamp precision is to the nanosecond. Scholtus, Van Dijk, and Frijns (2012) use ITCH data for State Street S&P 500 ETF. Scholtus and Van Dijk (2012) study three ETFs: State Street S&P 500 ETF, Powershares NASDAQ 100 ETF and iShares Russell 2000 ETF. Huh (2014) studies HFT liquidity provision, using constituents of the NASDAQ 100 index. Hasbrouck and Saar (2013) identify all domestic stocks in CRSP that are NASDAQ-listed, retaining the top 500 by market capitalization as of September 30, 2007.

Three studies analyse datasets of E-mini S&P 500 futures contracts[21]. E-minis are cash-settlement instruments with a notional value of $50 multiplied by the S&P 500 index and a minimum tick-size of 0.25 index points, the equivalent of 12.50$ per contract (Clark-Joseph 2013). According to the U.S. Securities and Exchange Commission (2014), E-minis generally lead price discovery on US equity markets as the most actively traded instrument in equity and equity related futures markets. Unlike other US equity instruments, however, E-minis are solely traded on the Chicago Mercantile Exchange (CME). The instrument is thus fully centralised on a single exchange. Transaction prices and quantities of E-minis as well as the aggregate

depth for each price level are observable through a public market-data feed (Clark-Joseph 2013). The CME provided the Computerized Trade Reconstruction (CTR) dataset to the CFTC, which, in turn, granted access to the authors of the aforementioned studies. The trade data include fields for price, number of contracts traded and the time of trading in units of seconds or milliseconds, depending on the month under examination (Baron, Brogaard, and Kirilenko 2012). The data are fully anonymised. Identities of traders are not released.

The NYSE Trade and Quote (TAQ) database contains intraday transaction data (trades and quotes) for all securities listed on NYSE, AMEX and NASDAQ National Market System (NMS). Hendershott, Jones, and Menkveld (2011) combine TAQ data with data retrieved from CRSP, focusing on NYSE common stocks. Gai, Yao, and Ye (2013), O'Hara, Yao, and Ye (2012) and Brogaard (2011) use TAQ data as a complement to the NASDAQ HFT dataset.

Other countries

Malinova, Park, and Riordan (2013) access a proprietary trader-level dataset provided by the TMX Group, owner and operator of the Toronto Stock Exchange (TSX). TSX data include all messages between brokers and exchange, including orders, cancellations, modifications and trade reports as well as whether the trade was buyer or seller initiated. Moreover, accounts are grouped by the clients the brokers cater for, which identify providers of Direct Market Access (DMA).

Another Canadian dataset is used by Brogaard, Garriot, and Pomeranets (2014), who obtained order messages and trade data from Alpha Alternative Trading System (ATS). Before its merger with the TMX Group, ATS accounted for over 20 percent of Canadian trading volume, making it the second-largest venue in Canada based on volume. The dataset contains all order messages and trade notifications sent to Alpha, including limit-order quotations, updates, fills and cancellations, time-stamped on the millisecond.

The Investment Industry Regulatory Organization of Canada (2012) uses a "regulatory feed" covering 11 Canadian equity trading venues, including TSX and Chi-X. The dataset contains both public and confidential regulatory information and includes data on all trade, order and quote messages for each venue as well as a User ID identifying, amongst others, individual DMA clients. The study period encompasses three calendar months (63 trading days), between August 1 and October 31, 2011. All listed securities included in the regulatory feed are included in the study, corresponding to 228 million trades and 9.86 billion order messages.

The Australian Securities & Investments Commission (2013) uses data from a "surveillance feed" provided by two Australian equity markets, the Australian Stock Exchange (ASX) and Chi-X. They focus on securities within the S&P/ASX 200, representing 95% of total equity turnover. Viljoen et al. (2014) use Thomson Reuters Tick History (TRTH) to obtain transaction and order information data on Share Price Index (SPI) 200 Futures contract traded on the Australian Securities Exchange. Their sample includes information on price, volume, buy and sell order ID and

---

[20] Brogaard (2010; 2011); Brogaard, Hendershott, and Riordan (2014); Carrion (2013); Gai, Yao, and Ye (2013); Gao and Mizrach (2011; 2013); Gerig (2012); Hasbrouck and Saar (2013); Hirschey (2013); Johnson, Van Johnson, Van Ness and Van Ness (2014); O'Hara, Yao, and Ye (2012); Tong (2013); Zhang (2013); and Zhang and Riordan (2011).

[21] Baron, Brogaard, and Kirilenko (2012), Clark-Joseph (2013) Kirilenko et al. (2011)

trade direction for all orders submitted between January and December 2009. The timestamp is to the millisecond.

Boehmer and Shankar (2014) obtained a dataset from the Indian National Stock Exchange (NSE) to study the effect of the introduction of several latency-improving infrastructure adjustments, such as DMA and colocation. The dataset contains each order-book message (entry, modification and cancellation) that arrives at the exchange and identifies the trader as AT or non-AT. Each trade contains information on the buy and sell order numbers. Around 1,400 stocks are traded on the NSE, the authors retain 150 in their sample, including the 50 constitutents of NSE's key index, the CNX Fifty, and 100 stocks selected from those traded in the derivatives segment.

Kang and Shin (2012) analyse a dataset containing KOSPI 200 futures contracts, which trade on the Korea Exchange (KRX). Its underlying index, the KOSPI 200, is a value-weighted index of 200 blue chips with a total market capitalization of 80 percent on the KRX. The authors' dataset includes all trades and order flows, time-stamped on a 10 millisecond-basis, allowing them to reconstruct the order book for each timestamp. Data fields include fields for price, quantity, type, order (or trade), sequence number and buy or sell indicator. The data identify individual trading accounts.

Bershova and Rakhlin (2013) use a cross-country dataset with data on British and Japanese equity markets, received via a large broker. The dataset, ranging from January to June 2010, contains aggregate daily volumes of stocks in the BE500 and Nikkei225 indices, "routed by HFT and LT [(i.e. long-term)] investors" (p. 6).

Cross-country Datasets

Two datasets were identified that use multi-venue samples across countries and regions. Aitken, Cummings, and Zhan (2014) use monthly data from 24 stock exchanges in 19 countries, including major stock markets in both developed countries and emerging markets.

Investigating algorithmic trading, Boehmer et al. (2012) construct a global sample encompassing 39 venues across developed countries and emerging markets. Thomson Reuters Tick History (TRTH) constitutes the main dataset and is combined with US intraday data from TAQ. The two are then merged with firm-level data from Datastream and CRSP. TRTH provides access to data feeds from various stock and derivatives exchanges, transmitted through the Reuters Integrated Data Network (IDN). The resulting sample includes 12,800 different common stocks. Data fields include intraday quotes and trades, which are time-stamped to the millisecond. Consequently, the authors do not have access to order book-level data.

Identification Strategies

From a research point of view, identifying traders as HFT is a means to an end, not an end in itself. HFTs are identified in order to commence the analysis, but the process of identifying them is seldom at the core of a study. Consequently, authors describe their identification methodology to a varying extent. This section summarises the most common approaches.

Articles are grouped into three categories, based on whether they use a direct, indirect or a mixed approach. Under the direct approach, HFTs are flagged via prior information on a firm's primary business from firm's websites, media or membership to certain trade associations. Identification via some common attribute, such as technological capacity in terms of low-latency infrastructure or computational power, is also considered to be a direct approach. The indirect approach, on the other hand, uses proxies to identify HFTs via common behaviour or strategies, such as low intraday and overnight inventories or high order-to-trade ratios. Mixed strategies are those which are composed of both direct and indirect strategies to identify HFTs.

Frequency of identification approaches in the HFT literature   C.2

Source: ESMA
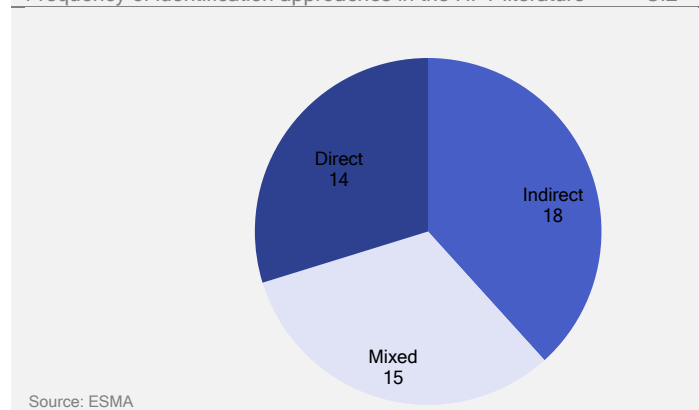
Direct Strategies

A number of authors use the direct approach to identify trading accounts as HFT, using prior information on the firm's core business. Benos and Sagade (2012), for example, use company websites and media reports to identify firms in the UK FSA's database whose core business includes HFT. In their sample, about 27 percent of trading value is characterised by HFT participation.

Brogaard, Hendershott, et al. (2014) and Alampieski and Lepone (2011; 2012) also use FSA data. The FSA and the trading venues (LSE, BATS, Chi-X) agreed which participants were HFTs based on their understanding of the participants' core business. This included knowledge regarding the use of algorithms and low-latency infrastructure. 52 participants were classified as HFT. In the dataset of Jarnecic & Snape (2014, p. 6), HFTs are similarly identified based on the LSE's "first-hand understanding of each client's predominant business". However, the dataset is aggregated on firm level, thus only allowing the distinction of order-book traffic between HFT and non-HFT firms but no identification of individual firms. Similarly, Hagstromer, Norden, and Zhang (2013) and Hagstromer and Norden (2013, 12) identify HFT traders "with the aid of NASDAQ OMX in-house expertise about member activity". Both identify 29 pure HFT and 22 hybrid firms, out of 100 firms in total.

Boehmer and Shankar (2014) obtained access to a dataset of the NSE of India, where algorithmic trading activity is identified by the venue, based on traders' use of algorithmic accounts. The dataset of Deutsche Boerse AG's

Xetra system similarly flags algorithmic trades, based on whether or not the account is registered in the Automated Trading Program (ATP), which offers reduced fees for automated traders. It is noteworthy here that Deutsche Boerse charges traders for executed trades, not submitted order book entries. Also on the NASDAQ OMXS, some authors take advantage of priory available information on the use of algorithmic trading accounts. Breckenfelder (2013), for example, identifies a subset of less than ten traders as HFT, each with about 10% market share on the Swedish exchange as well as significant market shares in the American market.[22]

Brogaard, Hagstromer, et al. (2013), who use a dataset similar to that of Breckenfelder (2013), also identify algorithmic traders via the use of algorithmic trading accounts, but separate them into different categories by speed, based on information on the use of colocation services. Bershova and Rakhlin (2013) use only ultra-low latency infrastructure as a proxy for HFTs. The downside of using technical capacity such as colocation as a proxy for HFT is stressed, amongst others, by Aitken, Cumming, and Zhan (2014). By identifying HFTs via the use of colocation services, they find that HFTs existence precedes that of colocation, not vice versa. This, they explain by HFTs requirement for high-speed transactions, which prompted exchanges to offer colocation services in the first place.

Indirect Strategies

Both Jovanovic and Menkveld (2010) and Menkveld (2013) study effects of one large HFT firm in their datasets. Ranking all broker IDs by the total number of messages sent to the exchange, they designate only the most active member as HFT.

Scholtus, Van Dijk, and Frijns (2012) identify HFTs via total message activity, i.e. the sum of all orders sent to the exchange, as well as several proxies based on fleeting orders. The concept of "fleeting orders" originates from Hasbrouck and Saar (2009) and is defined as an order that is added and removed from the order book within a given short period of time (s milliseconds). Using a threshold of 50ms to 100ms, they look at the percentage of fleeting orders to identify HFT activity. Moreover, they look at the number of fleeting orders that improve upon the best quote price as well as those that can be considered "missed opportunities" (Scholtus and Van Dijk (2012, p. 12) in the sense that they leave a worse order book when removed.

Kirilenko et al. (2011) define HFTs as a subset of intermediaries, i.e. short horizon traders that consistently run relatively low inventories. Two criteria are used to identify accounts who participate in a large number of transactions, while rarely accumulating significant net positions: (1) net holdings fluctuate within 5 percent of the end-of-day level; (2) the end-of-day net position (i.e. overnight inventory) must not exceed 5 percent of daily trading volume. Ordering the resulting list of intermediaries by their daily trading frequency, the top 7 percent are designated as HFTs.

Clark-Joseph (2013) uses the same characteristics to identify HFTs. Using daily volume to rank accounts whose overnight inventory changes by less than 6 percent of daily volume and whose maximal intraday inventory changes are less than 20 percent of daily volume, the top 30 accounts are designated HFT. Although these constitute less than 0.1 percent of the 41,778 accounts in his dataset, the designated HFTs participate in 46.7 percent of total trading volume and account for 31.9 percent of total message volume (entry, modification and cancellation of order book messages).

Kang and Shin (2012) classify trading accounts into three groups: HFTs, other ATs and ordinary traders. ATs are defined as traders submitting more than one thousand messages (entry, modification and cancellations) per day. A fraction of these ATs is designated HFT, based on two conditions: (1) overnight inventory must not exceed 3 percent of trading volume, and (2) fleeting orders are taken into account, by selecting accounts whose median cancellation time of limit order is smaller than 2 seconds.

Menkveld and Zoican (2014) adapt the strategies used by Kirilenko et al. (2011) as well. Accounts are identified as HFT if (1) their daily position change does not exceed 5 percent of volume and (2) the average difference between their minute-end and end-of-day positions does not exceed 1.5 percent of volume. The procedure identifies 5 out of 111 traders as HFT.[23]

Brogaard, Garriot, and Pomeranets (2014) identify HFTs who run both low intraday inventories and low overnight inventories. Individual traders are designated HFT for specific stocks if they (1) switch the sign of their trades (follow a buy by a sell order and vice versa) at least 33 percent of the time, and (2) if they hold no more than 20 percent of their daily volume overnight. They thus measure order book entries that "exhibit HFT qualities" (p. 11), rather than HFT firms.

Zhang (2010), studying quarterly data, extends the notion of zero overnight inventory, counting HFTs based on their end-of-quarter inventory

Baron, Brogaard, and Kirilenko (2012) use three thresholds a trader must cross in order to qualify as HFT: (1) trades must exceed a median of 5,000 contracts for all days the trader is active, (2) the median of overnight inventories, scaled by the total contracts the firm traded that day, must not exceed 5 percent, and (3) the maximum variation in intra-day inventories, scaled by total contracts, may not exceed 10 percent in the median over their sample. Using these proxies, they identify 65 out of 34,403 firms as HFT, which they classify further into "aggressive HFT" (initiates at least 60% of its trades), "passive HFT" (initiates less than 20% of its trades) and "mixed HFT" (neither "aggressive" nor "passive"). On average, HFTs account for about 54.4 percent of daily trading volume.

---

[22] Due to the confidentiality agreement with NASDAQ, the authors cannot release more accurate numbers on HFT activity.

[23] Since the TRTH dataset reveals the trader identity, Menkveld and Zoican (2014) can disclose the names of the five HFTs in their dataset: Citadel Securities, Spire Europe, International Algorithmic Trading GmbH, Getco Europe, and Nyenburgh Holding B.V.

Malinova, Park, and Riordan (2013) identify message-intense traders, which, according to their definition, include HFTs as well as message-intense agency algorithms, executed on behalf of institutional clients. Ranking trading accounts based on (1) their total number of messages and (2) their order-to-trade ratio, those in the top 5 percent of both measures are denoted "message-intense traders".

The Australian Securities & Investments Commission (2013) uses six measures that "relate strongly to the characteristics of [HFTs]" (p. 68): (1) order-to-trade ratios; (2) overnight inventory by total value traded, per security; (3) total daily value traded; (4) number of fast messages, defined as messages submitted within a 40ms window after an event (a) amendment or cancellation following previous action, or (b) a better-priced order is posted following a break in the market; (5) holding times, weighted by volume; and (6) at-best ratios, i.e. the sum of the number of orders placed at best price or at market divided by the total number of submitted orders. Subsequently, distributions are built around each metric and the distributions divided into quartiles. Traders are then scored based on their position in the distribution, receiving 4 points for being in the fourth quartile, three for the third quartile, and so on. The resulting index is then used to designate traders as HFT. For each day, the top 15 percent of the index are designated HFT for that trading day. Around 45 to 70 individual traders are thus identified on each day. Notably, the 10 largest designated HFTs account for 60 percent of value traded by HFTs (16 percent for the whole market), whereas the bottom 66 percent of designated HFTs accounts collectively account for only 1 percent of HFT trading. While the composition of the designated HFTs changes each day, the top 10 HFTs are always present.

To identify low-latency activity, Hasbrouck and Saar (2013) use a so called "strategic runs" measure. This measure is similar to the concept of "fleeting orders"; however, instead of focusing on the time between the entry and removal of an order, it focuses on the time between order cancellation and subsequent resubmission. To classify as low-latency, a new order of the same size and in the same direction has to be submitted within 100ms after cancellation of the previous order. About 60 percent of the cancellations in 2007 and 54 percent of the cancellations in 2008 could be linked this way. Although the threshold applied is 100ms, less than 10 percent exceed 40ms and 49 percent of the strategic runs identified arise from orders cancelled and resubmitted at 1ms or less,

Hendershott, Jones, and Menkveld (2011) similarly identify the overall activity of ATs, rather than individual accounts, focusing solely on total message traffic, i.e. order submissions, cancellations and trade reports. Their eventual measure of AT activity is the negative of total trading volume (in $1,000) divided by the number of electronic messages sent to the exchange. Boehmer, Fong, and Wu (2012), who do not have access to order-level messages, but only exchanges' best quotes and trades, use the same aggregate measure to identify AT activity. Specifically, they define ATs by the negative of the dollar volume associated with each message (defined as either a trade or a quote update). An increase in this measure reflects increases in AT activity. Viljoen et al. (2014) use the same measure to identify AT in their sample.

Mixed Strategies

The majority of studies using a mixture of direct and indirect strategies employ NASDAQ's HFT dataset. Although the identification is direct from an author's point of view in the sense that HFTs are pre-flagged in the dataset - the identification strategy for this dataset is a mixture of direct and indirect strategies to identify high frequency traders. Directly, traders are identified as HFT using "NASDAQ's knowledge of their customers" (Brogaard, Hendershott, and Riordan 2014, 5), including information on the firms' websites (Brogaard 2010). Indirectly, NASDAQ uses a number of proxies to identify traders on the basis of common attributes and trading patterns. Amongst the identification strategies used, Brogaard (2010) notes that NASDAQ directly identifies firms that (1) engage in proprietary trading, (2) make use of colocation services, and (3) engage in sponsored access provision. Indirectly, they use proxies to identify firms that (4) tend to switch between long and short net positions several times during the day, (5) have short time duration for orders, and 6) have high order-to-trade ratios. While the exact thresholds are unknown, 26 traders are thus identified as HFT. Carrion (2013) finds that HFTs participate in 68.3% of dollar trading volume. However, Brogaard, Hendershott, and Riordan (2014) find that they concentrate disproportionally on large liquid stocks. About 42 percent of trading volume in large stocks is accounted for by HFT, compared to just 11 percent in small illiquid stocks.

A limitation of NASDAQ's identification strategy is that it only includes what Brogaard, Hendershott, and Riordan (2014) refer to as "independent proprietary trading firms". It thus excludes large investment banks that operate HFT desks as well as small HFTs, who route their traffic through another member. Moreover, the dataset does not distinguish between different types of HFTs or non-HFTs, since the data are aggregated on the firm level. Another limitation is that the dataset only includes trade data, not individual messages sent to the order book of each individual stock.

The Investment Industry Regulatory Organization of Canada (2012) focuses on venue-provided information and order-to-trade ratios. Regarding the former, their dataset contains user IDs identifying, amongst others, DMA clients. The IDs are "assigned by a marketplace at the request of the participant" (p. 15). Order-to-trade ratios are computed by first de-trending the relationship between logged orders and trades, then using 1.25 standard deviations as a cut-off point on the distribution of this de-trended series to identify HFTs. The resulting group of about 316 HFTs has OTRs larger than approximately 11.2. Overall HFTs thus represent approximately 11 percent of trading accounts and account for roughly 42 percent of trading volume in each month.

Annex 2: Estimates of HFT activity for alternative HFT identification methods

In this section we present estimates of HFT activity for the following alternative HFT identification methods:

- Lifetime of orders approach based on the lifetime of orders at firm level
- Lifetime of orders of groups relative to the lifetime of orders on trading venues
- Message traffic

Lifetime of orders approach – firm level

Calculating the lifetime of orders at a firm level is a potential way to reduce the overestimation of HFT activity inherent to the calculation at group level. The intuition is that groups may be composed of both HFT and non-HFT firms. A calculation of lifetime of orders at group level may thus overstate the amount of HFT activity.

Indeed, calculating the lifetime of orders at firm rather than at group level reduces the amount of HFT activity in our sample. Overall, activity classified as HFT decreases from 43% to 37% in terms of value traded, from 49% to 42% in terms of the number of trades, and from 76% to 72% in terms of the number of orders (see C.11 in the main body of the paper).

The decrease in value traded and number of trades classified as HFT is not evenly distributed amongst different venues. A few markets (BATS, CHIX, TRQX) experience decreases of around 15pp, whereas there are only small decreases or even increases in HFT activity for the other venues, with the exception of MTAA.

An analysis of the underlying data shows that the decrease in HFT activity estimated at the firm level is largely driven by the exclusion of firms belonging to a small number of HFT groups. While at a group level their activity is correctly identified as HFT, they are falsely labelled non-HFT when the lifetime of orders is calculated at the firm level. Particularly on the venues where the decrease in activity was most pronounced, these groups appear to have predominantly used fill-or-kill orders. Fill-or-kill orders cannot be used for the calculation of order-lifetime, however, and the small number of their orders that remains for calculation is slower than our lifetime of orders threshold. Using the lifetime of orders approach at a participant level therefore introduces a large bias into the calculation results in our sample.

Overview of HFT activity - HFT flag and lifetime of orders at firm level						C.1
Trading venue	Value traded		Number of trades		Number of orders	
	HFT flag	Lifetime of orders	HFT flag	Lifetime of orders	HFT flag	Lifetime of orders
All venues	24	37	30	42	58	72
BATE	40	46	44	47	76	80
CHIX	40	44	40	45	59	79
MTAA	25	9	26	8	51	21
TRQX	34	48	35	47	73	79
XAMS	24	46	28	52	53	70
XBRU	18	50	23	52	38	56
XDUB	8	23	9	30	43	87
XETR	21	36	24	36	33	61
XLIS	11	42	17	46	31	54
XLON	21	32	26	34	44	57
XMCE*	0	25	0	25	0	43
XPAR	21	45	30	51	50	68

Note: Figures are weighted by value of trades (value traded), number of trades and number of orders, in %. For trades on UK stocks, value traded has been converted to EUR using end-of-day exchange rates.

BATE=BATS, CHIX=Chi-X MTAA= Borsa Italiana, TRQX=Turquoise, XAMS=NYSE Euronext Amsterdam, XBRU=NYSE Euronext Brussels, XDUB=Irish Stock Exchange, XETR=Deutsche Boerse AG, XLIS=NYSE Euronext Lisbon, XLON=London Stock Exchange, XMCE=Mercado Continuo Español, XPAR=NYSE Euronext Paris.

*No HFT firms were direct members of XMCE during the observation period. Therefore no HFT activity is reported for XMCE under the HFT flag approach.

Source: ESMA.

Lifetime of orders - group vs firm approach			C.2
Trading venue	Value traded		
	HFT flag	Lifetime of orders group approach	Lifetime of orders firm approach
All venues	24	43	37
BATE	40	60	46
CHIX	40	56	44
MTAA	25	20	9
TRQX	34	63	48
XAMS	24	48	46
XBRU	18	48	50
XDUB	8	19	23
XETR	21	35	36
XLIS	11	40	42
XLON	21	32	32
XMCE*	0	32	25
XPAR	21	45	45

Note: Figures are weighted by value of trades (Value traded in %). For trades on UK stocks, value traded has been converted in EUR using end-of-day exchange rates.

*No HFT firms were direct members of XMCE during the observation period. Therefore no HFT activity is reported for XMCE under the HFT flag approach.

Source: ESMA.

Lifetime of orders relative to the lifetime of orders on trading venues

A lifetime of orders approach based on a relative threshold is one of the options for HFT identification in ESMA's Technical Advice to the European Commission.

In practice, each trading venue could, for example, calculate the median lifetime of orders and determine cases in which the median lifetime of individual members' orders falls below that for the market as a whole. Where this is the case, a firm would be identified as an HFT.

The results for a number of thresholds are presented, to show the sensitivity of this approach to their values. The thresholds presented are based on the order duration at venue-level, ranging from the first decile (10% shortest-lived orders) to the fifth decile (median order-lifetime for market). The tables below show the percentage of HFT activity based on value traded.

In a first instance, only the activity of a firm in stocks where the lifetime of orders is below the threshold lifetime of the trading venue is characterised as HFT. In that sense, this step of the approach is comparable to the approach used in the headline results in this paper for the lifetime of orders approach. Results for HFT activity vary significantly between the different thresholds (C.3). For example, activity by investment banks amounted to 61% of the total value traded in the sample. Using the 30th percentile, the value traded of those banks that would be considered as HFT would be 5% of the total value traded in the sample.

For all participants, using the first decile of order lifetime at venue-level as threshold, 1 percent of value traded is identified as HFT activity compared to 30 percent of value traded for a threshold at the fifth decile.

Lifetime of orders - relative approach. Activity classified as HFT under different thresholds		C.3				
Direct approch	Total value traded	Threshold value				
		10th Percen-tile	20th Percen-tile	30th Percen-tile	40th Percen-tile	50th Percen-tile
HFT	24	1	2	4	7	15
IB	61	0	2	5	7	12
Other	15	0	0	0	1	2
All	100	1	4	9	15	30

Note: % of value traded to total value traded. Total value traded considers all the activity by members using only their classification by the Direct approach. Columns under "Threshold value" indicate the % of value traded to total value traded in the sample that is classified as HFT according to the relative lifetime approach before using the upgrade rule, i.e. using a stock by stock approach.

Source: ESMA.

Secondly, results are presented using an "upgrade rule". Under the upgrade rule, participants identified as HFT for one particular stock on a trading venue are classified as HFT for all activity on that trading venue. Such a concept could be applied in a regulatory context, where e.g. certain organisational or record-keeping requirements are linked to being identified as a HFT firm.

With the upgrade rule (C.4), results for HFT activity vary between 16 percent of value traded for a threshold at the first decile of order lifetimes on the trading venue and 78 percent of value traded for a threshold at the fifth decile of order lifetimes on the trading venue.

Lifetime of orders - relative approach. Activity classified as HFT under different thresholds. Upgrade rule		C.4				
Direct approch	Total value traded	Threshold value				
		10th Percen-tile	20th Percen-tile	30th Percen-tile	40th Percen-tile	50th Percen-tile
HFT	24	9	13	18	21	22
IB	61	7	16	26	34	47
Other	15	0	2	3	4	8
All	100	16	31	47	59	78

Note: % of value traded to total value traded. Total value traded considers all the activity by members using only their classification by the Direct approach. Columns under "Threshold value" indicate the % of value traded to total value traded in the sample that is classified as HFT according to the relative lifetime approach after using the upgrade rule, i.e. considering as HFT any activity of a firm that was considered as such in at least a stock.

Source: ESMA.

Another interesting aspect is to analyse how many participants classified as HFT firms under the HFT flag approach would be identified under the order lifetime approach with a relative threshold. In our sample a total of 181 participants are labelled as HFT firms and are active in May 2013 (C.5). 61 out of these 181 participants would be captured as HFT with the first decile as threshold. When using the fifth decile, this increases to 153 out of 181 participants.

Lifetime of orders relative approach. Number of firms classified as HFT under different thresholds		C.5				
Direct approch	Total Population	Threshold value				
		10th Percen-tile	20th Percen-tile	30th Percen-tile	40th Percen-tile	50th Percen-tile
HFT	181	61	84	118	145	153
IB	319	40	75	123	162	221
Other	711	21	57	92	131	191
All	1211	122	216	333	438	565

Note: Total population column contains the number of firms in each of the available categories according to the Direct approach. Columns under "Threshold value" indicate the number of firms that are classified as HFT according to the relative lifetime approach.

Source: ESMA.

Message traffic

Under the German HFT Act, part of the identification process for HFT firms is based on message traffic. Message traffic is also one of the options for HFT identification in ESMA's Technical Advice to the European Commission.

We have calculated the message traffic of firms on a stock-by-stock basis and present results for both the number of firms and the percentage of trading activity classified as HFT in our sample. A number of message traffic thresholds are presented, since the key question is above which threshold trading activity should be characterised as HFT. The message thresholds are calculated for the trading hours of the trading venues and range from two messages per second to one message every 10 seconds.[24]

---

[24] Differences in daily trading times for trading venues are taken into account.

As our dataset is a sample of stocks traded, we have not calculated a message traffic threshold at a trading venue level. Such a threshold (e.g. 75,000 messages per day) would likely lead to more activity captured as HFT on large trading venues due to the higher number of stocks traded relative to small venues.

In a first instance, only the activity of participants in particular stocks where they exceed the message threshold, is characterised as HFT. Thus, this step of the approach is comparable to the approach used in the headline results in this paper for the lifetime of orders approach. Results (C.6) for HFT activity vary significantly between 5 percent of value traded for a threshold of two messages per second to 48 percent of value traded for a threshold of one message every 10 seconds (0.1 messages per second).

messages per second in our sample of stocks. For a threshold of 0.1 messages per second this increases to 75 out of 181 participants.

Message traffic approach. Number of firms classified as HFT under different thresholds		Threshold value							C.8
Direct approach	Total Popu-lation	2	1.5	1	0.75	0.5	0.25	0.1	
HFT	181	16	24	30	37	40	57	75	
IB	319	5	7	16	20	30	59	96	
Other	711	0	1	1	1	2	7	12	
All	1211	21	32	47	58	72	123	183	

Note: Total population column contains the number of firms in each of the available categories according to the Direct approach. Columns under "Threshold value" indicate the number of firms that are classified as HFT according to the message traffic approach.

Source: ESMA.

Message traffic approach. Activity classified as HFT under different thresholds		Threshold value							C.6
Direct approach	Total value traded	2	1.5	1	0.75	0.5	0.25	0.1	
HFT	24	3	6	8	10	12	15	17	
IB	61	2	3	7	8	11	17	30	
Other	15	0	0	0	0	0	1	1	
All	100	5	9	15	19	24	32	48	

Note: % of value traded to total value traded. Total value traded considers all the activity by members using only their classification by the Direct approach. Columns under "Threshold value" indicate the % of value traded to total value traded in the sample that is classified as HFT according to the message traffic approach before using the upgrade rule, i.e. using a stock by stock approach.

Source: ESMA.

Secondly, results are again presented using an "upgrade rule". This means where a participant is classified as HFT in one stock on a trading venue, all its activity on that trading venue is classified as HFT.

With the upgrade rule, results (C.7) for HFT activity vary between 13 percent of value traded for a threshold of two messages per second and 63 percent for a threshold of 0.1 messages per second.

Message traffic approach. Activity classified as HFT under different thresholds. Upgrade rule		Threshold value							C.7
Direct approach	Total Value Traded	2	1.5	1	0.75	0.5	0.25	0.1	
HFT	24	9	11	13	16	16	18	20	
IB	61	4	6	12	13	17	28	42	
Other	15	0	0	0	0	0	1	2	
All	100	13	18	25	30	33	47	63	

Note: % of value traded to total value traded. Total value traded considers all the activity by members using only their classification by the Direct approach. Columns under "Threshold value" indicate the % of value traded to total value traded in the sample that is classified as HFT according to the message traffic approach after using the upgrade rule, i.e. considering as HFT any activity of a firm that was considered as such in at least a stock.

Source: ESMA.

As above, a total of 181 active participants in our sample are labelled as HFT firms using the HFT flag approach (C.8). Only 16 out of these 181 participants would be captured as HFT with a message traffic threshold of two

Annex 3: Stratified approach used to identify market participants

For each trading venue, the list of all market members during the reporting period was requested, including internal member ID, name of the member and Bank Identification Code (BIC).

For each individual account on a venue, an anonymised ESMA ID was created. When a market member had two accounts on the same venue (same name and BIC), those two accounts had two different ESMA IDs but the same ESMA Account ID.

Since BICs were not always provided and were not always unique to a firm, an approach based on names was used to map market members across trading venues. Some expressions were excluded (like Limited, Ltd etc.) and the resulting names were compared. When there was a match, the different market members share the same Group ID. When only parts of the name matched (for example Bank XYZ Europe and Bank XYZ London), the market members were given different Group IDs but the same Master ID.

Number of IDs				C.1
Type	ESMA ID	Account ID	Group ID	Master ID
All	1211	773	394	348
HFT	181	96	20	20
IB	319	187	57	25
Other	711	490	317	303

Note: Number of IDs in the sample
Source: ESMA.

Annex 4: Sampling procedure used for establishing the sample of stocks

The sample of 100 stocks used in our analysis has been established using a stratified approach. We had to restrict the number of stocks in our analysis due to storage and processing issues and to take into account the cost of providing the data for the trading venues. On the other hand, we had to ensure that relevant segments of European stock markets are appropriately represented in the sample (i.e. stocks of all participating countries as well as a diversified sample of large, medium and small-cap stocks).

Below we describe the stratification process we used in order to select our sample.

As a first step, all listed stocks included in the Thomson Reuters Datastream list of companies in European markets index were included (2579 stocks from 30 countries). The sample was reduced to 1619 stocks by including only the stocks traded in the nine EU countries for which data were collected (BE, DE, ES, FR, IE, IT, NL, PT and UK).

For the stratification process, September 2012 data have been used. Data on market value are based on Thomson Reuters Datastream. Data on value traded of stocks are based on Bloomberg.

Market value of EU stocks considered							C.1
Country	Number	Mean	Min	Q1	Median	Q3	Max
BE	90	2,298	64	154	372	892	105,875
DE	248	4,869	117	297	744	4,001	93,075
ES	248	4,228	144	260	631	2,298	68,411
FR	40	1,131	1	52	386	1,336	7,080
IE	159	2,214	71	170	480	1,389	63,853
IT	119	3,716	0	31	343	2,164	98,566
NL	50	899	4	20	67	579	10,025
PT	122	3,358	5	65	282	1,826	60,837
UK	543	4,182	0	300	725	2,243	139,507
Total	1,619	3,723	0	213	591	2,125	139,507

Note: Market value in EUR mn as of September 2012.
Sources: Thomson Reuters Datastream, ESMA.

The minimal number of stocks per country was chosen to be five and the maximum to be 16. Within those constraints, the final number of stocks selected per country was determined in order to be proportional to the relative sizes of financial markets.

Trading concentration for each stock has been estimated as the ratio of value traded in the trading venue where the stock has its primary listing to total value traded in all venues. Fragmentation across trading venues was also computed for each stock, as the ratio of value traded in the trading venue where the stock has its primary listing to consolidated value traded for each stock.

The universe was further reduced from 1619 to 1317 stocks by imposing that the stocks should fulfill one of the following conditions:
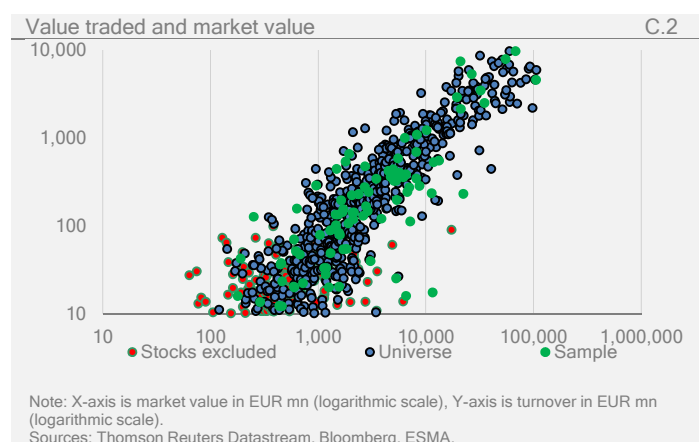
i) Value traded in September 2012 is higher than EUR 100mn; or
ii) Value traded is higher than EUR 1mn and fragmentation is lower than 0.9 (value traded on the trading venue where the company has its primary listing is less than 90% of total value traded).
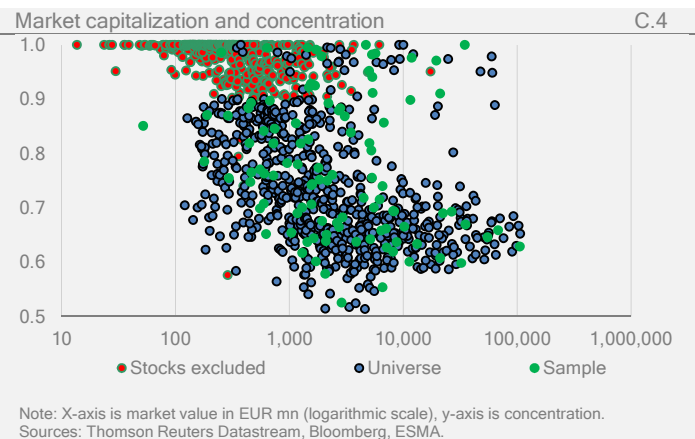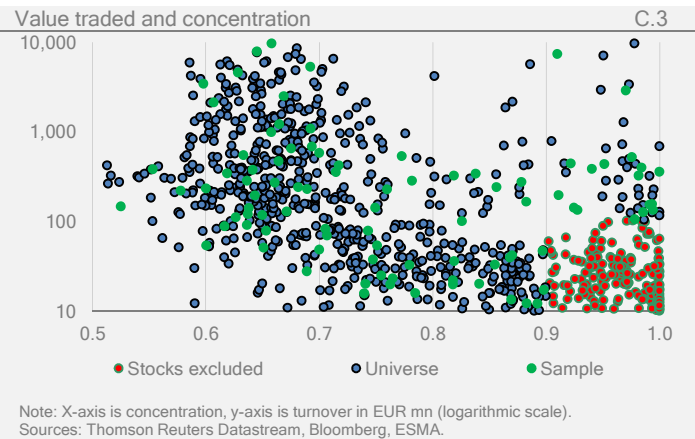
For the complete sample, four quartiles based on market capitalization were generated. After that, since market capitalization and value traded are highly correlated, value traded data were regressed on market capitalization to create two categories of stocks using the estimation error: stocks with high value traded given their market capitalization and stocks with low value traded. The same process was used to derive two groups of stocks with respect to fragmentation, using a Tobit model.

This process provided us with a total of 16 categories. Each stock in the population was classified into one of these 16 categories. All stocks were then assigned a randomly generated number. For each country, the stock in each of the 16 categories with the highest random number was chosen to be the candidate for the respective category. For countries where 16 stocks were required in our sample, all of the 16 candidates were chosen to be part of our sample. For countries where less than 16 stocks were required in our sample, we selected the relevant number of categories with a view to have a sample of stocks which is as representative as possible of the population of stocks in that country.

Charts C.2 to C.4 show the relationship between value traded and market value, value traded and the concentration of trading, as well as market capitalization and concentration of trading. It also shows which of the stocks out of the possible universe were included in our final sample as well as which stocks originally supplied were not included in the universe used for sampling.

Table C.5 provides an overview of the characteristics of our final sample.

Value traded and market value   C.2

Note: X-axis is market value in EUR mn (logarithmic scale), Y-axis is turnover in EUR mn (logarithmic scale).
Sources: Thomson Reuters Datastream, Bloomberg, ESMA.

Value traded and concentration   C.3

Note: X-axis is concentration, y-axis is turnover in EUR mn (logarithmic scale).
Sources: Thomson Reuters Datastream, Bloomberg, ESMA.

Market capitalization and concentration   C.4

Note: X-axis is market value in EUR mn (logarithmic scale), y-axis is concentration.
Sources: Thomson Reuters Datastream, Bloomberg, ESMA.

Sample stocks statistics   C.5

Country	Value traded (EUR mn)			Market Cap (EUR bn)			Fragmentation Index		
	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
All sample	33.7	611.3	<0.1	8.7	122	<0.1	0.4	0.7	0
BE	45.7	357.1	0.3	24.3	122	0.8	0.4	0.7	<0.1
DE	37.1	611.3	<0.1	8.2	73	<0.1	0.3	0.7	0
ES	42.8	526	2.6	9.6	41.8	0.7	0.3	0.5	<0.1
FR	34.8	497.2	<0.1	7.5	58	0.1	0.4	0.7	0
IE	5.3	184.7	<0.1	3.6	8.1	<0.1	0.2	0.6	0
IT	33.1	300.7	<0.1	6.5	28.2	0.3	0.2	0.6	0
NL	37.3	350.5	0.3	7.7	51	0.4	0.4	0.7	0.1
PT	17.2	143.1	<0.1	5.3	11.4	2	0.3	0.6	0
UK	29.2	290.2	0.1	8.5	71.2	0.4	0.5	0.7	0.1

Note: Note: Monthly average, minimum and maximum for May 2013. For the fragmentation index a value of 0 indicates no fragmentation (all trading is on one venue), whereas higher values indicate that trading is fragmented across several trading venues.
Source: ESMA.

Annex 5: Presentation of the database

The database includes anonymised data on market participants and their characteristics (market maker, colocation, etc.) as well as data on trades and orders collected from trading venues through the National Competent Authorities.

Characteristics of market members

Trading venues were asked to provide the following information on their members:

- Anonymised ID,
- Status of the firm (Credit Institution/Investment firm/Non-MiFID firm),
- The list of stocks in our sample for which the member was designated market maker/liquidity provider and
- Flags for colocation, provision of Direct Market Access and Sponsored Access.

Extract from Member ID table						C.1
Anon ID	Status	Market maker	Colocation	DMA	SA	MIC
XYZ	IF	ISIN1	Y	N	Y	TV1

Note: Mock-up data
Source: ESMA.

Trade data

Trade data were collected for all stocks in our sample during the reporting period (May 2013). The main fields cover:

- The anonymised ID of the buyer and seller,
- The corresponding order ID with date and timestamp (usually up to the millisecond),
- The capacity of the buyer and seller (proprietary/agent trading),
- A direction field (aggressive order) and
- The price, currency, size and ISIN code of the stock.

Other fields were included to identify auctions and cross trades.

Extract from trade table						C.2
Buyer ID	Seller ID	Stock	Price	Size	Direction	Timestamp
XYZI	ABCD	ISIN1	10.21	150	B	10:23:18.800

Note: Mock-up data
Source: ESMA.

Order data

Order data include the anonymised ID of the member, along with the characteristics of the order:

- Type (limit order, market order, etc.),
- Validity (good till day, fill or kill, etc.),
- Action (New/Modified/Cancelled),
- Price and size of the order and

- Additional indicators (visible and hidden quantities for Iceberg orders, pegged orders, stop price, agent-based or proprietary trading).

Extract from order table						C.3
Member ID	Order type	Validity	Action	Buy/Sell	Direction	Timestamp
XYZI	Limit	GTD	New	Buy	B	10:23:12.457

Note: Mock-up data
Source: ESMA.