

Hossain, Shahadat

Article

High-Frequency Trading (HFT) and market quality research: An evaluation of the alternative HFT proxies

Journal of Risk and Financial Management

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Hossain, Shahadat (2022) : High-Frequency Trading (HFT) and market quality research: An evaluation of the alternative HFT proxies, Journal of Risk and Financial Management, ISSN 1911-8074, MDPI, Basel, Vol. 15, Iss. 2, pp. 1-31, <https://doi.org/10.3390/jrfm15020054>

This Version is available at:

<https://hdl.handle.net/10419/258778>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Article

High-Frequency Trading (HFT) and Market Quality Research: An Evaluation of the Alternative HFT Proxies

Shahadat Hossain ^{1,2} 

¹ Department of Finance, University of Chittagong, Chittagong 4331, Bangladesh; shahadat@cu.ac.bd

² Department of Economics and Statistics, University of Naples Federico II, 80126 Naples, Italy; shahadat.hossain@unina.it

Abstract: We examine the soundness of high-frequency trading (HFT) proxies that are widely defined on the limit order book (LOB) information. We use a unique TRTH (Thomson Reuters Tick History) millisecond time-stamped intraday trades and quotes dataset enriched with 10 levels of LOB depth messages for 149 highly fragmented LSE listed stocks for the period 2005 to 2016. We explore a sharp uptrend in HFT activities and accompanying improvement in market liquidity in the European market. We show that alternative HFT proxies built on LOB are not equally powerful. The HFT proxy defined on the five best LOB prices (the mid point of a typical limit order book) provides a better HFT identification than the one popularly defined on the first best prices (BBO). We suggest that picking the LOB information beyond a certain level (e.g., the best five prices) of market depth in developing HFT proxy is counterintuitive. Evidence indicates that high-frequency traders (HFTs) participate in both competitive (narrow) and passive (wider) quoting as a market making strategy; however, they do not participate in passive quoting excessively.

Keywords: market microstructure; high-frequency trading (HFT); HFT proxy; limit order book (LOB); best bid and offer prices (BBO); liquidity



Citation: Hossain, Shahadat. 2022. High-Frequency Trading (HFT) and Market Quality Research: An Evaluation of the Alternative HFT Proxies. *Journal of Risk and Financial Management* 15: 54. <https://doi.org/10.3390/jrfm15020054>

Academic Editor: Thanasis Stengos

Received: 20 December 2021

Accepted: 19 January 2022

Published: 25 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For the last two decades, the advent of sophisticated computing technology has been changing the financial market structure unprecedentedly. Machines are gradually occupying the places for which human interaction was necessary. High-frequency trading (HFT) is the latest and one of the important technological inclusions in the modern trading platforms. The contribution of HFT in equity trading is significantly large across developed financial markets. For example, in the USA, it was approximately 52% of total equity trading in 2018 (Zaharudin et al. 2021). HFT is extraordinarily fast and the fastest speed has increased the effectiveness of HFT strategies in the present-day fragmented trading environment (Baldauf and Mollner 2021). It is now a regulatory concern all over the world regarding the market impact of HFT. Addressing these issues, market microstructure research in HFT has been trending for the last decade and an increasing body of empirical literature has shown contradictory evidence.

Surprisingly, to date, the consensus on the definition of HFT is quite low. Nonetheless, the differences in research evidence and the public, media, and regulatory perception regarding the impact of HFT are huge (Gomber et al. 2011). For empirical study, HFT identification is a precondition. The limited and costly access to HFT data limits the scope of choices for the researchers, and studies to rely on some imperfect HFT identification strategies. Proprietary trade and quote data extracted from the limit order book (LOB) are one of the broadly accessible data sources in HFT studies (Ben Ammar and Hellara 2021; Boehmer et al. 2018; Conrad et al. 2015; Friederich and Payne 2015; Frino et al. 2017; Hendershott et al. 2011; Leone and Kwabi 2019). Generally, those datasets are supplied without direct identification or HFT flag, and users must rely on some definitions or proxies

for HFT identification. The choice of proxies are many folds (Biais and Foucault 2014; Bouveret et al. 2014) and a single data source does not support them all. Proprietary LOB datasets which are commonly used for HFT research are also source-dependant, and the choice of proxies are mostly constrained by the fields and depth levels to which an user gains access.

On top of that, managing LOB data is always challenging because of its voluminous aspect, computational burden, and required time and resource commitment for its management. Therefore, compromises in the choices of necessary data fields is quite common. For example, many studies (Boehmer et al. 2021; Hendershott et al. 2011) do not use the LOB information beyond the best market depth level or BBO, either for avoiding the related complexities in data management or not having the appropriate data access. Yet the evidence (AMF 2017; Hendershott and Riordan 2013) on ATs/HFTs' participation at the different LOB levels shows a participation rate of 52% to 70.8% at the BBO and marketable orders, and 64% to 79.3% behind the BBO and nonmarketable limit orders respectively. It implies that a proxy defined on the first best level of limit order book messages does not incorporate the HFT activities behind that point and could create a number of caveats which concerns the validity of the proxy. On the other hand, expanding the data extraction to many levels inside an order book poses the challenge of data handling. Moreover, the use of diverse HFT proxies in the absence of a proper standard is creating many concerns. Zaharudin et al. (2021) put the right apprehension of using distinct proxies across HFT research: "The lack of a uniform identification for HFT leads to problems, such as research complications, that lead to somewhat conflicting conclusions as to the effect of HFT on equity markets in general and the market microstructure in particular."

In this paper, we attempt to evaluate the soundness of a set of alternative HFT proxies that are dependent on proprietary LOB data and frequently used in the HFT and related studies. The primary aim is to find an appropriate level of limit order books' depth until which the extraction of information helps the best identification of HFT footprints, and simultaneously minimizes the issues related to the insignificant use of voluminous data. In doing so, we gain access to one of the richest millisecond time-stamped TRTH (Thomson and Reuters Tick History) intraday trades and quotes dataset enriched with 10 depth levels messages of LOB covering the period 2005 to 2016 for a group of 149 LSE (London Stock Exchange) listed large capitalized stocks in Europe. We develop all proxies available in the relevant literature and supported by our datasets, and evaluate them in two phases: (i) firstly, we discuss their descriptive properties and issues that have already been reported in the literature; (ii) afterwards, we employ them in a broadly used market microstructure empirical setting to understand their effectiveness in identifying the HFT activities.

A first-hand analysis shows that Hendershott et al. (2011)'s proxy (dollar volume per electronic message times (-1)) does not appropriately fit in the European equity market setting for the sample period under our consideration¹. Another proxy, order to trade ratio, also has the same limitation. For determining the impact of choices of different level of limit order book information on HFT proxies, we take two extremes and a midpoint of the limit order book market depths, and define three alternative HFT proxies on them. We evaluate each of the three proxies by employing them in our empirical models. The analyses show that the HFT proxy defined on the first best bid and offer prices (BBO) fails to include the footprints that HFTs leave on a typical limit order book through their passive (nonmarketable) quoting. The HFT proxy constructed by taking the other extreme of limit order book depth, i.e., the first 10 best bid and ask prices, does not also show the improvement with regard to explaining the changes in market quality parameters. On the contrary, the HFT proxy defined on the mid point or the first five levels of market depth information provides the best HFT identification and reasonable explanation for the changes in market quality parameters over the past. The analyses also confirm that market fragmentation has a significant role in understanding the HFT behaviour.

This study contributes to the existing body of empirical market microstructure literature in general and high-frequency trading literature in particular in at least two ways.

Firstly, to our knowledge, this is the first study of its kind to evaluate the alternative high-frequency trading proxies built on limit order book information. Secondly, it provides evidence regarding the high-frequency trading intensity and market liquidity changes in the European equity market covering one of the longest periods, using an original dataset. The rest of the paper is organized as follows. Section 2 gives a short accounts of the HFT market microstructure in the European market, Section 3 describes the literature. Section 4 provides the description of data, variable definition, and empirical specification. Section 6 reports and discusses the results, and Section 6 concludes the study.

2. Market Microstructure Setting

2.1. High-Frequency Trading

As a new field in the Algorithmic Trading (AT) domain, the literature and definition on HFT are in their infancy (Gomber et al. 2011). Aldridge (2013) highlights that the computing technology plays the biggest role in HFT business. HFT identifies every small change in the limit order book quote updates and enables them to move faster. A fairly volatile financial market that ensures adequate changes in price to exceed transaction costs and provides ample space so that traders can quickly take in and out from market positions, befits HFTs' need. There are many papers (Aldridge 2013; Biais and Foucault 2014; Gomber et al. 2011; O'Hara 2015; The Netherlands Authority for the Financial Markets 2010; Zaharudin et al. 2021) that have made attempt to paint a descriptive picture of HFT and have marked a distinction between the working feature of HFT and AT. It reveals that both HFT and AT share some common features but they work on different purpose and capacity. 'Market making' is one of the main strategies of HFT. Additionally, the domain of HFT strategies includes electronic liquidity provision, statistical and latency arbitrage, liquidity detection, and short-term momentum. Sometimes HFTs are registered as the designated liquidity provider in addition to market making. The salient features of HFTs include (Gomber et al. 2011): very high number of order generation, fast order revision or cancellation, proprietary trading, generating profit from market making, maintaining a zero or minimal inventory, fast holding revision, using low-latency technology, taking position in highly liquid instruments, and so on. For a detail discussion on HFT and AT, we refer the reader to Aldridge (2013); Gomber et al. (2011); The Netherlands Authority for the Financial Markets (2010). Figure 1 shows the domain of HFT.

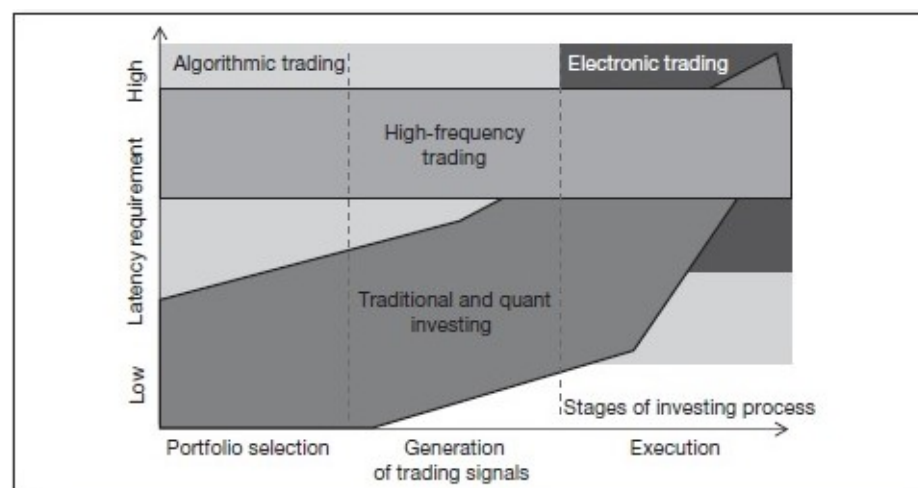


Figure 1. High-Frequency Trading Domain (source: Aldridge 2013).

2.2. European Market Structure

The adoption of MiFID in 2007 (MiFID II in 2018) has been proliferating the trading venues across the European equity market. One of the vital changes it has brought into effect was abolishing the monopoly power of traditional exchanges in trading securities and liberalizing them into many trading platforms such as regulated markets (RMS),

multilateral trading facilities (MTFs), and systematic internaliser (SIs). These platforms have different market structures and reporting systems that are defined under MiFID directives².

In broad terms, RMs and MTFs operate in a similar fashion, provide an electronic platform for users to transact orders multilaterally. These trading venues generally match orders on a non-discretionary basis according to pre-defined rules that establish price and time priority for submitted orders. RMs and MTFs are required to publish pre-trade quotes and report details of executed trades to the market (CFA Institute 2011). Both RMs and MTFs are allowed to organize primary listing. However, RMs facilitate the listing of regulated instruments and MTFs do the same for the unregulated ones. A firm chooses on which RM to list, and once listed, MTFs may decide to organize trading for that firm as well. SIs are investment firms that internalise order flow to deal on their own account on an organised, frequent, and systematic basis. Trades executed through SIs are reported as the over-the-counter (OTC) trades.

The large RMs in the European equity market include the LSE Group (operator of the London Stock Exchange and Borsa Italiana), NYSE Euronext (which operates exchanges in France, Belgium, the Netherlands, Portugal, and the United Kingdom), and Deutsche Borse Group (operator of the Frankfurt Exchange and the Xetra trading system). The LSE runs electronic order books on which buy and sell orders are continuously matched from the open to the close according to the price-time priority rules. Automated trading sessions start at 8:00 and close at 16:30 in local time. As a supply side response, LSE has been investing heavily in technology against growing HFT demands for low-latency³ tradings for the last two decades. The implementation of the Millennium trading platform has improved its latency to 113 microseconds which were 600 milliseconds before the year 2000 (Linton and Mahmoodzadeh 2018). Beside RMs, the main MTFs are CHIX, BATS, and Turquoise. These exchanges are well equipped with modern latency-based technologies, and have become main rivals of primary exchanges like LSE.

2.3. Limit Order Book (LOB)

The understanding of a limit order book structure is fundamental for developing HFT proxies. Every new information arrival keeps the LOB updating. It may occur due to new order, order modification, order cancellation, or trade execution, etc. The limit order and market order are two basic order types that investors generate in the system according to their needs of liquidity. Investors can design orders differently, according to their needs, while submitting them to an LOB. A successful limit order execution depends on the finding of a counterparty with the opposite needs and therefore sometimes remains unexecuted. A market order is executed immediately if it finds an outstanding limit order on the other side of the LOB. All limit orders are arranged on both sides of the LOB according to their price and time priority. The lowest available price on the ask/offer side (sell limit order) at a particular point of time is known as the best offer price. All offer prices other than the best one are queued behind the best price according to their time and price priority. The highest available price on the bid side (buy limit order) at a particular point of time is known as the best bid price. All bid prices other than the highest one are also queued behind the best price according to their time and price priority. The best bid and offer prices, BBO, are the best available prices on both sides of a limit order book at a particular time. The construction of a hypothetical LOB is illustrated in Figure 2. The depth level of an LOB refers to the number of offer and bid prices in a queue. For example, 'n'th-level depth refers to the 'n'th best bid and ask prices. We refer Table A1 in the Appendix A for a real-time LOB update up to the 10 levels. HFT studies generally establish a link between the intensity of HFTs' participation in the limit order book and the frequency or speed at which quotes are updated there (Conrad et al. 2003; Hendershott et al. 2011).

Market sell order of 200 (or limit sell with price < 74.42)			Market buy order of 900 (or limit buy with price > 75.74)		
Price	Size	Time	Price	Size	Time
74.42	500 100	11:49:39	74.48	300	11:49:35
74.41	100	11:46:55	74.48	500	11:49:40
74.36	400	11:48:30	75.74	100	08:25:17
74.36	400	11:48:32	76.00	150	08:02:02
74.00	13	10:56:00	76.77	20	07:01:01
73.75	5100	11:28:02	77.00	100	09:15:00
72.98	5100	10:56:99	77.06	200	10:14:11
72.15	120	08:01:39	77.35	1000	08:01:39
72.11	20	07:01:01	77.82	20	07:01:01
72.03	20	07:01:01	78.00	300	08:02:00
72.00	Because of the buy market order, the bid-ask spread widens from 74.48 – 74.42 = 0.06 to 76.00 – 74.42 = 1.58				9:30:04
71.59	The market order has “consumed” liquidity.				8:01:32
71.11					9:30:04
71.00					7:01:01
70.35					8:01:35
70.11	20	07:01:01	80.00	350	09:15:00

Figure 2. Example of Limit Order Book (Source: Foucault et al. 2013).

2.4. High-Frequency Data

High-frequency data are the records of the real-time LOB updates. Every change in a LOB, also known as ‘tick’, shows the continuous update of a limit order book’s records as a result of the arrival of new limit orders (buy and sell), revision or modification of the existing limit orders, cancellation of the existing limit orders, or the entry of new market orders. “Ticks” are sequentially collected and recorded with a proper timestamp in producing the highest-frequency data. A timestamp records the date and time at which a particular event of quote updates occurs. The timestamps show the time and date at which either the exchange or the broker-dealer released the quotes, or the time at which the automated trading system received the quotes at the exchange, or when the third-party distributor of the ‘tick’ data stamp the time and date at its own facility after receiving the real-time data feeds from the exchange. Each record of high-frequency data is comprised of many fields providing the records for security identifier, exchange identifier, time region, transaction types, trades or quotes volume, prices, qualifiers for showing the different market conditions, and many more. We refer the reader to Aldridge (2013) for a good coverage on high-frequency data.

The importance of high-frequency data in studying complex microstructure of modern financial markets is well established. Nowadays, market microstructure studies are using high-frequency data more frequently than ever before. Although high-frequency data have concentrated on the major financial markets, many recent studies on less developed or emerging markets report their high-frequency data properties (Markellos et al. 2003). Modern investors search for profitable opportunity across electronic markets through their intraday high-frequency trading windows (The Netherlands Authority for the Financial Markets 2016). Electronic exchanges are connected with each other through the traders. Kollias et al. (2013) draw the inference from relevant studies: “intraday data help reveal a more accurate picture of how markets and market agents react and adapt to changes and exogenous shocks...the more detailed account and the information contained in high frequency data allow more reliable inferences and conclusions to be drawn vis-à-vis daily data”.

3. Relevant Literature

HFT identification strategy is not only critical but a central issue in empirical studies. The empirical HFT literature is divided into many strands of which HFT is considered a subset. A researcher attempting an empirical HFT study has limited choices, and depends on either one or both of the approaches for identification: (i) to use an exchange provided HFT-flagged dataset (direct method); or (ii) to define a proxy which tracks the footprint of HFT (indirect method). Conrad et al. (2015) has pointed out the limitation of using the

exchange-identified HFT data. Generally, exchange houses select samples according to their own criteria, and this may not be free from possible conflict of interest which prevails among the users of HFT flagged data, HFT firms, and trading platforms. Other than that, sample firms appear to be large and specialized in HFT often operate in several exchanges across countries. There are various reasons why there could be a non-random distribution of trades across trading venues due to their heterogeneity in liquidity, fee structure, etc. Therefore, sometimes, drawing inferences from these datasets may not reflect the true HFT behavior. In contrast, a well-defined proxy tracks the predominant market making nature of HFT through posting and renewing quotes.

3.1. HFT Studies Using Direct Method

HFT-flagged data provide marks internally on every order originated by the high-frequency traders (HFTs) with a distinct flag. Many studies have used a pre-identified dataset. The NASDAQ dataset is the famous flagged dataset that has been used in many HFT papers such as Carrion (2013); Brogaard et al. (2014); O'Hara et al. (2014); Brogaard and Garriott (2018). The dataset identifies a subset of HFTs for the years 2008, 2009 and a week of 2010 only for a stratified sample of stock. The dataset provides the HFT flags for marking the demanders and suppliers of liquidity. It includes 26 HFT firms identified by NASDAQ manually, which make the dataset unique.

The other pre-identified datasets used in the prior studies are mostly study-specific and diverse. Brogaard et al. (2014) used the HFT flagged FSA data set measuring the HFT activities in the UK equities asset of particular HFT firms from 5 November, 2007 to 5 August 2011. Brogaard and Garriott (2018) used the data from one of the alternative trading platform of Canada with complete order-book information. The dataset provides the identification for direct-access members and enables the identification of HFT firms using inventory and trading behaviour. It also provides the sequence by which the members get entered in the market. Hendershott and Riordan (2013) used HFT flagged data from the Deutsche Boerse. The dataset contains all orders generated from AT in 30 DAX stocks (the leading index for German stock market) between 1 January and 18 January 2008, for a total of 13 trading days.

3.2. HFT Studies Using Indirect Method

Unflagged HFT data are trades and quote data supplied from the exchange houses without pre-identification or flag. The users need to develop some sorts of proxies for identifying the HFT activities. Ideally, the definition of an HFT proxy target one or more HFT identifying features, and develop the measure accordingly. The seminal study of Hendershott et al. (2011) used the 'number of electronic messages per \$100 trading volume' as their proxy is defined only on the BBO level information for the period 2001–2005. They targeted the 'speed' as an identifying feature of HFT. There are also other features upon which many studies have relied. For example, Kirilenko et al. (2017) used the 'daily net position' with the idea that HFTs usually perform high-volume trades and carry low intraday and overnight positions. Hasbrouck and Saar (2013) developed the proxy 'Strategic Runs' of linked messages where the proxy exploits the particular order sending and cancelling pattern of low-latency traders. Moreover, quote updates (Conrad et al. 2015), message-to-trade ratios (Friederich and Payne 2015; Frino et al. 2017), the duration or 'life time of orders' (Bouveret et al. 2014), and HFT strategies (Hagströmer and Nordén 2013), etc., are proxies which have been defined on one or more identifying features of HFT.

4. Data, Variables and Measures, and Methods

4.1. Data

LSE is the largest RM in Europe for listing the highest number of large market capitalized stocks from Stoxx 800, a benchmark index constituting the largest 800 market capitalized stock in Europe. It is widely believed and demonstrated by previous studies that large-caps stocks are highly liquid, attract more HFT, and are extremely fragmented

(AMF 2017; Boehmer et al. 2021; Gresse 2017; Hendershott et al. 2011). We analyse the list for STOXX 800 at the end of 2016 to know its composition (See Figure A1). It shows that that top 50% of the stocks in the list are from only three primary trading venues, the London Stock Exchange (LSE), Deutsche Boerse (Xetra), and Euronext Paris, of which LSE listed stocks are more than 50%. It also shows the market share of both primary and alternative lit trading venues in European equity markets. Among the trading venues, CHIX, BATS, and Turquoise facilitate most of the lit trading besides the primary platforms. After a long trial, we finalize our sample that includes large market capitalized stocks of LSE which are considerably liquid, cross-listed in MTFs during 2005–2016. Table 1 lists all securities by their RICs (Reuters instrument code) and available data points (number of days) which are included in the sample.

Table 1. The list of all LSE listed sample stocks. The table lists all 149 LSE listed securities (by their RICs) included in the sample for the period 2005–2016 with number of trading days coverage.

RIC	Days	RIC	Days	RIC	Days	RIC	Days	RIC	Days	RIC	Days
AAL	3028	BYG	2892	GOG	2991	LAND	3028	RB	3029	SPX	2973
ABF	3027	CCL	3028	GPOR	2989	LGEN	3028	RBS	3028	SRP	3017
ADN	2898	CLLN	2989	GRG	2977	LLOY	3029	REL	3028	SSE	3028
AHT	2898	CNA	3028	GRI	2972	LSE	3022	REX	2900	STAN	3028
ANTO	3024	CNE	3029	GSK	3028	MCRO	2855	RIO	3028	SVS	2968
ARM	2944	COB	3024	HIK	2816	MGGT	3010	RMV	2722	SVT	3028
ASHM	2572	CPG	3028	HLMA	2993	MKS	3028	ROR	2973	SXS	2977
AV	3027	CPI	3026	HMSO	3022	MNDI	2399	RR	3028	TATE	3028
AVV	2814	CRDA	2934	HSBA	3027	MONY	2382	RRS	2968	TCG	2408
AZN	3028	DEB	2691	HSV	2975	MPI	2861	RSA	3027	TLW	3026
BAB	2917	DGE	3028	HSX	2972	MRW	3028	RTO	3028	TPK	3004
BARC	3028	DLN	2503	ICP	2980	MTO	2987	SAB	2966	TSCO	3028
BATS	3028	DOM	2764	IGG	2942	NEX	3021	SBRY	3028	TW	2397
BBA	3020	DRX	2786	IHG	3028	NG	2886	SDR	3023	UBM	3027
BDEV	3028	DTY	2874	III	3028	NXT	3028	SGC	3016	ULE	2980
BLND	3027	ECM	3021	IMI	3024	OML	3023	SGE	3028	ULVR	3027
BLT	3027	ELM	2882	INCH	3021	PFC	2838	SGRO	2427	UTG	2956
BNZL	3027	EMG	3028	INF	2869	PFG	3023	SHB	2971	UU	3028
BOY	2986	EXPN	2581	INVP	2991	PNN	3009	SHP	3028	VOD	3029
BP	3029	EZJ	3025	ITRK	3008	PRU	3027	SL	2646	WEIR	2992
BRBY	3011	FDSA	2429	ITV	3028	PSN	3026	SMDs	2997	WG	2999
BT	3028	FGP	3025	JLT	2994	PSON	3028	SMIN	3028	WMH	3027
BVIC	2789	GFS	3022	JMAT	3027	PZC	2956	SMWH	3007	WPP	3021
BVS	3005	GKN	3028	KGF	3029	QQ	2748	SN	3028	WTB	3017
BWY	3019	GNK	2992	KIE	2915	RAT	2958	SNR	2858	Total days	439,583

The primary source of our data is Thomson Reuters Tick History (TRTH), a product of the Securities Industry Research Centre of Asia-Pacific (SIRCA), which is compiled from the Global Thomson Reuters exchange feeds. Two resilient London-based recording devices provide the millisecond timestamps to each recorded message. The primary analysis of TRTH data structure reveals that time synchronizations of trades and respective quote messages are not uniform across trading venues. TRTH provides better time synchronizations between quotes and trades for the trading venues which are physically closer to the IDN Collection LAN in London (e.g., LSE, CHIX, Bats, Turquoise) than for those which are not (e.g., Deutsche Boerse (Xetra), Euronext Paris). This issue raises some real challenges in determining trades and quotes based measure of transaction costs, which is particularly true for the effective spread. Considering the TRTH time synchronization issue, we narrow down the sample choices only to the UK-based LSE-listed stocks. To address the fragmented environment of these stocks appropriately, we select CHIX, BATS, and Turquoise as their alternative venue counterparts. These four trading venues facilitated around 99%

of lit tradings during the period 2014–2016 for the stocks that are primarily listed in LSE. The trades and quotes data are available from TRTH since 1996 for most of the primary trading venues, which for alternative trading venues in MiFID zone started to be available from mid of 2008. Among the 220 primarily selected stocks from the LSE, TRTH provides data support only for 204 stocks. At this point, a primary analysis shows that among those 204 securities, some are not compatible for further analysis due to reasons such as delisting, takeovers, or mergers with other firms or liquidation at some point or not having enough data coverage for all four trading venues for unknown reasons, and we reduce the sample finally to 149 stocks.

TRTH supplies intraday quotes and trades records through two main files—the Time and Sales (TS) and the Market Depth (MD). Two resilient London-based recording devices provide the millisecond time stamp to each recorded message. The time and sales file provides transaction records and the best quote updates, and the market depth file comes with the queue of bid and ask limit prices and respective quantities, as displayed in the limit order book. The records in market depth can be extracted to 25 best limit prices (based upon their availability) of which we limit our data up to the 10 best prices since no records are found beyond that level. The full limit order book provides important information about the depth and spread beyond the BBO, which essentially affects the trading decisions of market participants. As far as we are aware, the sample we construct here should provide the longest coverage of data with its highest granularity used in any HFT research. We also rely on the Thomson Reuters’s Datastream for some other data such as control variables which are not made available from TRTH.

4.2. Variables and Measures

4.2.1. High-Frequency Trading

We define all HFT proxies available from the literature, and that are supported by our datasets. Therefore, we can not construct proxies that require a direct link-up of every limit order book update with an order ID. We can track intraday millisecond records of the trades and quotes data for every update in the LOB. For an update in any fields of its 10 level depths, we see the change and track them. The updates are generally caused by any of the following reasons: (i) order execution, (ii) arrival of new limit order, (iii) quote cancellation, and (iv) quote modification. We then aggregate and divide them by the number of minutes allocated for each daily automated trading session (8.00 to 16.30) to get the daily quote update speed per minute.

We use the method used in [Hendershott et al. \(2011\)](#) for defining the daily quote update speed per minute known as the electronic message rate (hft_{it}). The electronic message rate measures the intensity of the LOB quote updates according to the HFT speed. The electronic message rate (hft_{it}) is

$$hft_{k_{it}} = \frac{qupdate_{it}}{T},$$

where $qupdate_{it}$ is the aggregate daily quote update up to the k level depth of the LOB for stock i on day t , and T is the length of a daily trading sessions (in minutes).

A recent regulatory study report ([AMF 2017](#)) shows that HFTs actively participate beyond the best limit prices. The average market share of HFT in the BBO, the two best limit prices, and the three best limit prices are 70.8%, 77.3%, and 79.3%, respectively. We use the definition of daily electronic message rate per minute for measuring three alternative HFT proxies (hft_bbo_{it} , hft_5bo_{it} , hft_3bo_{it}) based on three different market depth level (k) information. hft_bbo_{it} , hft_5bo_{it} , hft_3bo_{it} measure the quote update until the BBO, 5 best bid and ask prices, and 10 best bid and ask prices, respectively.

Our fourth alternative measure for HFT is [Hendershott et al. \(2011\)](#)'s proxy where electronic message rate is adjusted for trading volume ($algo_trad_k_{it}$). We define

$$algo_trad_k_{it} = \frac{value_{it}}{qupdate_{it}} * 100 * (-1),$$

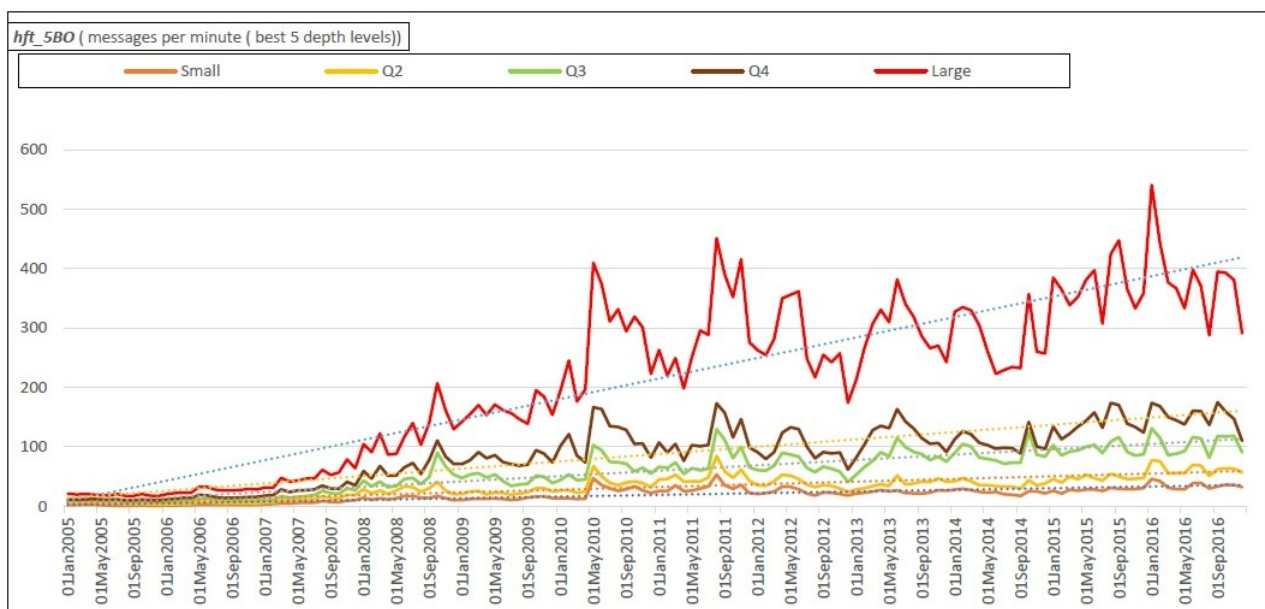
where $value_{it}$ is the value of the trading volume of stock i on day t , $qupdate_{it}$ is the aggregate daily quote update up to the k level depth of the LOB for stock i on day t . We also define two alternative measures of the same metric, $algo_trad_(10bo)_{it}$ and $algo_trad_(5bo)_{it}$ for the 10 best levels and 5 best levels order book updates, respectively.

Our last proxy is the order-to-trade ratio ($ord_to_trd_{it}$) and is defined as below,

$$ord_to_trd_{it} = \frac{qupdate_{it}}{ntrades_{it}},$$

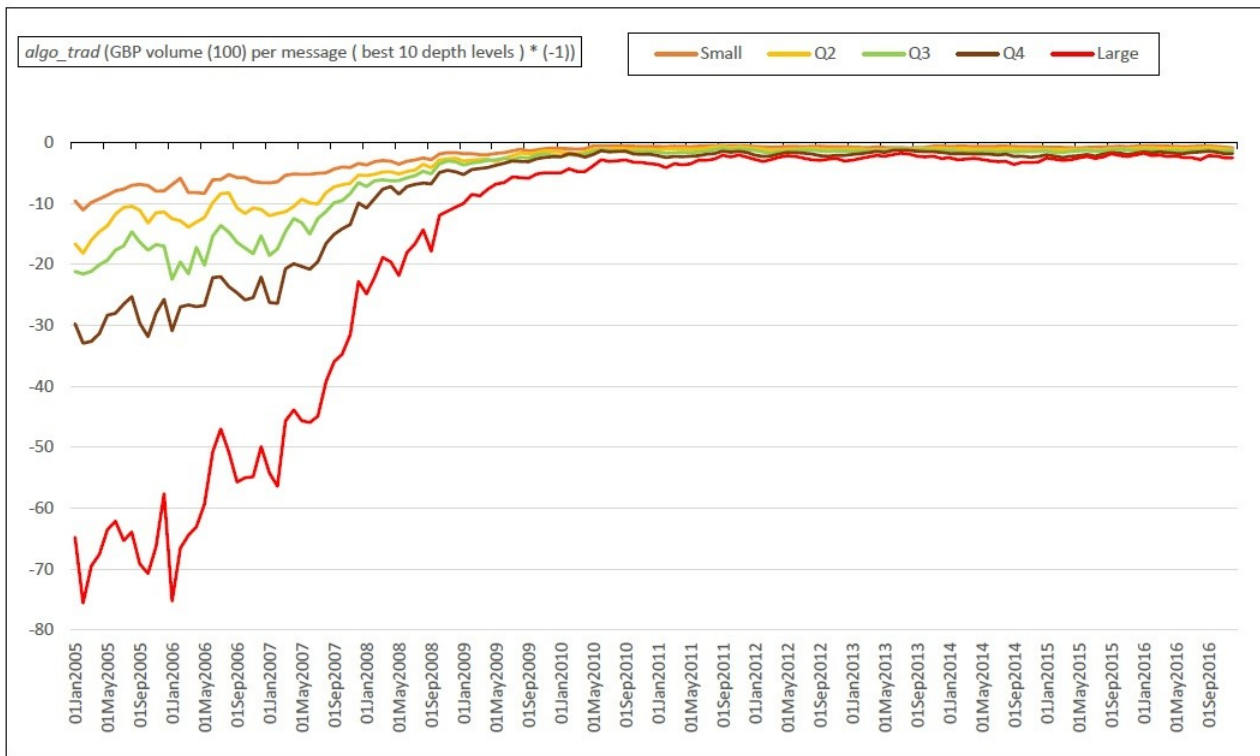
where $qupdate_{it}$ is the aggregate daily quote update up to the 10-level depth of the LOB for stock i on day t , and $ntrades_{it}$ is the number of executed trade of stock i on day t .

The post-MiFID period is prominently characterized by the proliferation of low-latency based modern trading venues, through which markets have been experiencing a large influx of HFT investment. A substantial upward trend of the message trafficking during this period can be seen in [Figure 3a](#). [Table 2](#) shows the descriptive statistics of all HFT measures where 149 stocks are divided into 5 quintiles, based on market capitalization. It shows that HFT intensity increased over the sample period remarkably ([Figure 3](#)). The average message traffic rate hft_10bo in 2005 starts at 9 messages per minute, and rises to 176 messages in 2016. All other HFT proxies also show a similar rise in HFT activity during this period. It can also be seen that HFT activities are more intense in large stocks. The average message speed for the proxy developed on 5 best bid and offer price (hft_5bo) is more than double that for the proxy constructed on the BBO (hft_bbo). It can also be seen that HFT intensity demonstrated in the measure hft_10bo is not that intense, as captured by the hft_5bo .

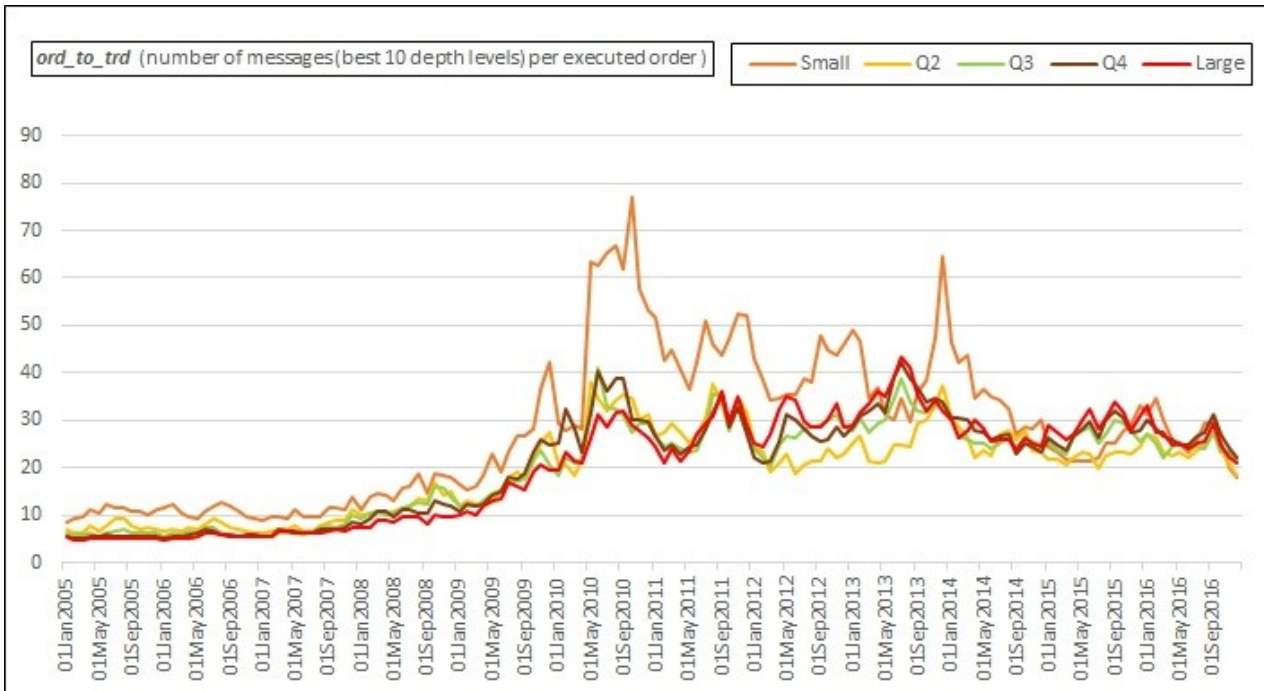


(a)

Figure 3. Cont.



(b)



(c)

Figure 3. High-frequency Trading Measures. These figures depict: (a) the traffic of electronic messages per minute (b) [Hendershott et al. \(2011\)](#)'s proxy of HFT and (c) order-to-trade ratio.

Table 2. Descriptive statistics for HFT proxies. The table reports the descriptive statistics for all HFT proxies. The sample comprises 149 stocks divided into 5 equal quintiles based on market capitalization. Table 1 shows the data coverage in details.

HFT Proxy	Description	Statistics	All	Small	Q2	Q3	Q4	Large
<i>hft_10bo</i>	messages per minute (best 10 depth levels)	Mean	100.36	23.77	38.97	72.39	103.80	266.31
		Median	47.62	17.95	30.57	56.44	90.99	171.75
		StdDev	157.77	23.22	38.25	67.59	84.81	271.25
<i>hft_5bo</i>	messages per minute (best 5 depth levels)	Mean	83.72	19.35	32.61	60.80	89.36	219.40
		Median	41.10	14.42	26.09	49.34	80.49	149.97
		StdDev	126.10	19.37	30.18	52.20	68.56	213.53
<i>hft_bbo</i>	messages per minute (BBO)	Mean	39.32	9.57	15.67	28.33	43.50	100.91
		Median	20.38	7.48	12.87	23.91	40.04	76.56
		StdDev	54.32	8.98	13.09	21.76	30.81	88.30
<i>ord_to_trd</i>	number of messages for the 10 best levels per executed order (order-to-trade ratio)	Mean	22.02	28.05	19.51	20.45	21.18	20.73
		Median	19.06	19.97	17.28	18.46	20.40	19.83
		StdDev	21.25	36.67	15.59	14.76	13.57	13.84
<i>algo_trad(10bo)</i>	GBP volume (100) per message (best 10 depth levels) time (−1)	Mean	−6.93	−2.24	−3.73	−5.14	−7.51	−16.21
		Median	−1.96	−0.88	−1.37	−1.77	−2.38	−3.60
		StdDev	15.51	4.19	5.87	8.42	10.93	29.12
<i>algo_trad(5bo)</i>	GBP volume (100) per message (best 5 depth levels) time (−1)	Mean	−7.54	−2.80	−4.17	−5.53	−8.04	−17.37
		Median	−2.33	−1.10	−1.64	−2.05	−2.72	−4.19
		StdDev	16.22	5.30	6.32	8.78	11.36	30.28
observations (stock * day)			439,583	90,046	88,374	87,239	86,786	87,138

4.2.2. Market Fragmentation

For developing a sound empirical strategy, we incorporate the market fragmentation in our setup. The idea is that technology and regulation drive both HFT, and fragmentation and they are closely related (Menkveld 2016; Upson and Van Ness 2017). The impact of order fragmentation on liquidity is well documented in many studies, e.g., Degryse et al. (2015); Gresse (2017); O’Hara and Ye (2011); Aitken et al. (2017), etc. We refer to the study Mishra and Zhao (2021) for a more comprehensive discussion regarding the link between HFTs and fragmented equity markets.

We use the Herfindahl–Hirschman Index (*HHI*)—the most commonly used definition of market concentration in the literature, as the proxy for order fragmentation across markets. We define the proxy for stock level volume fragmentation, *HHItrd_{it}*, as

$$HHItrd_{it} = \frac{1}{\sum_j^n v_{ij}},$$

where *v_{ij}* is the square of the trading volume share on venue *j* among *n*, for security *i* at day *t*. This is a normalized measure that ranges from 1 to *n*, where 1 stands for no fragmentation or full concentration in the primary venue, and *n* for evenly distributed order flow across *n* exchanges. Since our study considers one primary venue (LSE) and three alternative exchanges, the range is $1 \leq HHItrd_{it} \leq 4$. The trading volume of all stocks, on average, has fragmented across the venues over the past. We find that the average *HHItrd* has increased from 1 in 2005 to 2.89 in 2016 (Figure 4). It may be noted that the market was concentrated before 2007, and the proxy takes the value 1 accordingly.

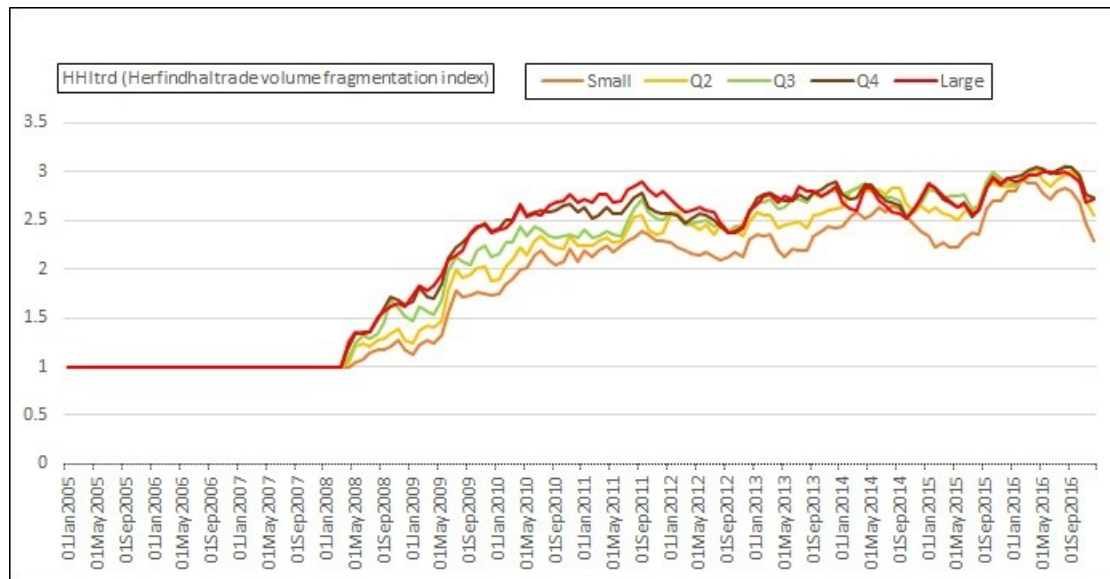


Figure 4. The trading volume fragmentation proxy (*HHItrd*).

4.2.3. Liquidity Measures

We take only the liquidity⁴ dimension of market quality measures. Market microstructure literature classifies liquidity into several dimensions, such as transaction costs, quantity, time, and so on. We use the quoted half-spread (*spread_bps_{it}*) and the effective half-spread (*espread_{it}*) as measures of the transaction cost dimensions of liquidity. To address the quantity dimension of liquidity, the quoted depth (*depth_k_{it}*) is measured at and beyond the BBO.

For the *tth* intraday quote in stock *i*, quoted half-spreads in basis points (*spread_bps_{it}*) are defined as

$$spread_bps_{it} = ((ask_{it} - bid_{it}) / mp_{it}) * 10000,$$

where *ask_{it}* is the best quoted ask price, *bid_{it}* is the best quoted bid price, and *mp_{it}* is the quote midpoint at BBO calculated as $(ask_{it} + bid_{it}) / 2$. Finally, we aggregate all intraday *spread_bps_{it}* into a time weighted daily measure. For interpretation, the lower the quoted spread, the higher the liquidity and vice versa.

For the *tth* intraday trade in stock *i*, the effective half-spread in bps, *espread_{it}*, is

$$espread_{it} = (d_{it} * (p_{it} - mp_{it}) / mp_{it}) * 10000,$$

where *d_{it}* is an indicator variable that equals +1 if the *kth* intraday trade is a liquidity demander’s buy and −1 if the *kth* trade is a liquidity demander’s sell, *p_{it}* is the trade price, and *mp_{it}* is the quote midpoint prevailing at the time of the *kth* trade. Finally, all intraday effective spreads are weighted by their trade volume to convert them into a daily measure. For clarification, a narrower effective spread is associated with a higher liquidity and vice versa.

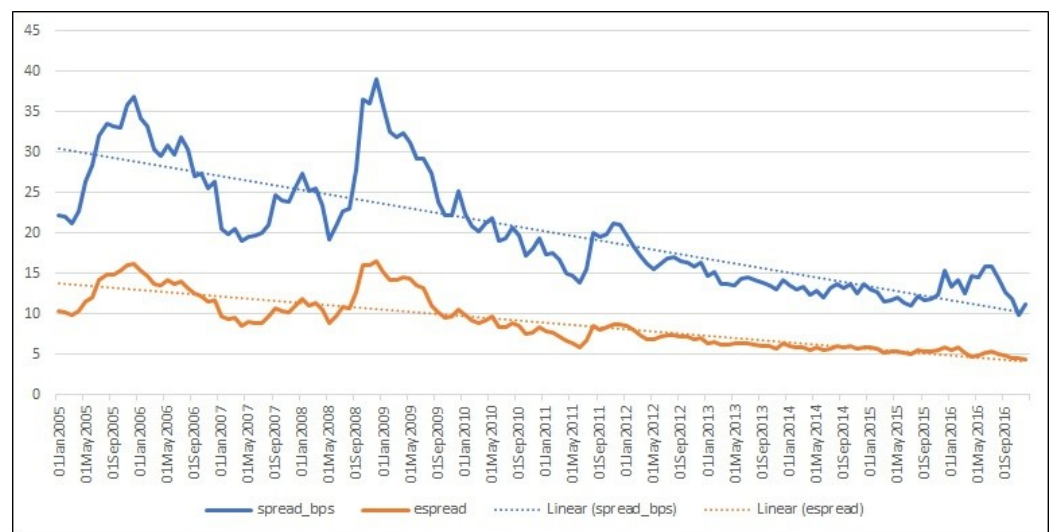
Generally, empirical studies use the readily available trade signing approaches found in the literature, e.g., Lee and Ready (1991). The signing of trade using any of these methods comes with a high cost of inaccuracy due to the fact that exchange platforms and data providers usually do not follow a uniform data synchronization system. In contrast to common practices, we develop unique algorithms that are capable of signing trade precisely in TRTH European data structure. The algorithms match every trade price with the immediate prevailing quotes, both bid and ask, and define *kth* trade as liquidity demander’s buy if it matches quoted ask price, and as liquidity demander’s sell if it matches quoted bid price.⁵

The average quoted depth can be decomposed into offer depth (the specified quantity that a liquidity supplier is willing to sell at ask price), and bid depth (the specified quantity that a liquidity supplier is willing to buy at the bid price). The intraday quoted depth for stock ‘i’ at time ‘t’ is

$$depth_{k_{it}} = (ask\ depth * ask\ price + bid\ depth * bid\ price) * 0.5,$$

where $depth_{k_{it}}$, refers to the average offer and bid quantity available at k level of market depth at time t , for stock i . We define quoted depth for two values of k , the first one, $depth_{bbo_{it}}$, is based on BBO level market depth, and the other one, $depth_{3bo_{it}}$, is based on the first three best market depth levels. Finally, we aggregate all intraday depth into a daily measure by taking simple average. For interpretation, the larger the depth size, the higher the liquidity supply at depth level ‘ k ’.

During the period 2005–2006, both quoted spreads and effective spreads reveal substantial improvement of liquidity in the market (see Figure 5a). For instance, overall quoted spreads decreased more than one-half and dropped to 13.5 bps (approx.) at the end of 2016 compared to 30 bps (approx.) in 2005. The effective spreads also decreased to approx. 4 bps compared to approx. 10.7 bps in 2005. Figure 5b shows that both depth measures of liquidity declined over the years significantly. It would seem surprising if one interprets this movement as liquidity degradation. We rather argue that the decline in depth might be attributable to the narrowed depth as market makers have less incentive to offer a larger depth that is usually done for a wider spread. The rising HFT intensity during the sample period is also linked to the decreasing trade sizes. HFT is generally performed in small lots, and it prominently uses slice-and-dice strategy in executing a large order (Aitken et al. 2017; Gresse 2017; Hendershott et al. 2011).



(a)

Figure 5. Cont.



(b)

Figure 5. Liquidity measures. These figures depict three liquidity measures: (a) quoted half-spreads and effective half-spreads; (b) quoted depth at the BBO and 3 best levels. Dotted lines shows the linear trend of these variables.

4.2.4. Alternative HFT Proxies: Do They All Fit?

One of the early used HFT/AT proxies in market microstructure literature is $algo_trad_{it}$ which was originally developed and used in the study [Hendershott et al. \(2011\)](#). The proxy is defined as the negative of trading volume (in hundreds of dollars) divided by the number of messages. The idea of adjusting the message rate by the trading volume originated from the fact that the sample of NYSE used in the study for the period 2001–2005 was associated with an upward trend of the trading volume. However, this adjustment makes this proxy difficult (or misleading) for interpretation in some cases. Generally, the lower the ratio is (the higher the absolute proxy value), the higher the HFT intensity is. It can be seen that the interpretation does not hold when the same ratio is used to compare HFT intensities across stocks or groups.

Let us imagine that there are two different stocks, 1 and 2. For stock 1, the value of $|algo_trad_{it}|$ is v_1 and that for stock 2 is v_2 where $v_1 < v_2$. Let us imagine again, for any stock, the value of $|algo_trad_{it}|$ at time t_1 and t_2 are v_{t_1} and v_{t_2} , respectively. For showing a rise in the HFT intensity over time, the condition is $v_{t_2} < v_{t_1}$. On the contrary, for the same comparison across stocks (let us say intensity of HFT in i is higher than that of j), the condition should be $v_i > v_j$, i.e., the opposite. This difficulty of comparison makes the proxy less attractive in different scenarios. For example, we refer to [Figure 3b](#) to illustrate the theoretical interpretation issue. An intensive growing tendency of the electronic message rate (a decreasing trend of the absolute value of the proxy) can be noted across all quintiles over the sample period, and the pattern confirms the condition we set, $v_{t_2} < v_{t_1}$, accordingly. Again, in the same figure, the quintiles (large) showing the highest HFT intensity lie at the bottom of the groups (the lowest absolute value of the proxy), i.e., $v_{large} > v_{small}$. This explanation confirms that the cross-sectional and time-series comparison is somewhat dubious for the proxy $algo_trad_{it}$.

Another HFT proxy, the order-to-trade ratio ($ord_to_trad_{it}$) is a popular measure used by the regulators in many countries like Canada, Australia, Germany for controlling HFTs. Studies such as [Frino et al. \(2017\)](#); [Friederich and Payne \(2015\)](#) have used order-to-trade ratio. Among others, [Bouveret et al. \(2014\)](#) evaluated its pros and cons as a proxy for HFT or AT activities, and finally concluded: “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". Apart from that, the issue regarding the dubious comparison that we have discussed against *algo_trad* is equally applicable for this measure. We refer to Figure 3c for a similar comparison.

'Overnight inventory position' (Jovanovic and Menkveld 2015; Kirilenko et al. 2017), 'strategic runs' (Hasbrouck and Saar 2013), 'order life time' (Bouveret et al. 2014) are some study specific proxies which development have made the choice set of proxies larger. However, to the best of our knowledge, we do not find any studies using them afterward. The idea of using intraday overnight inventory position as a proxy is motivated by the fact that HFTs reportedly carry very low intraday and overnight inventory position, and an HFT is distinguished from a non-HFT after comparing the day closing inventory position level. The difficult part of this strategy is that one should be able to identify and distinguish all traders with some classes of flags supplied in the data set. Unfortunately, most of the time, the supply of datasets of this sort is very limited. The procedures described in the Hasbrouck and Saar (2013) for constructing strategic runs are also not easy and almost difficult to construct for data unavailability. The 'order life time' is defined as the time elapsed before the order is modified and cancelled. Generally, HFTs send orders with a shorter lifetime. To trace the order life time, an order id is an essential field in the HFT dataset. Nonetheless, the datasets which are enriched with 'order id' but do not have enough information of the limit order book's depth are also not sufficient for replicating the proxy.

We argue that among the proxies referred here, the electronic message rate (hft_{it}) should be the best choice for being chosen as an HFT proxy. This is because the definition is simple, and the interpretation is quite straightforward. It also highlights the main technological aspects of the HFT revolution for the last two decades, i.e., speed. For this reason, it is seen that many papers have used this definition and constructed the proxy on the limit order book information. However, the concern is that most of the studies have relied only on the first-level market depth (BBO) in defining their proxies. The reasons for this might be that BBO level data are highly available, less costly, and require less time and resource commitment for processing. Conversely, evidence coming from regulatory case studies like AMF (2017) shows that HFT participates in quotes beyond the BBO. The AMF studied the activity of the leading HFTs on Euronext Paris focusing the order book liquidity provided by the leading HFTs. It reports that the presence of the HFTs at the BBO, two best price limits and three best price limits were 70.8%, 77.3%, and 79.3%, respectively, compared to the total liquidity provided in the market. We see that the participation of HFTs at the deeper of the limit order book are quite high but we do not have evidence until which level they do participate.

4.3. Methods

To determine the level of HFT participation in the limit order book, we take three benchmark levels—BBO, the five best prices, and the ten best prices, and define three proxies— hft_bbo_{it} , hft_5bo_{it} , and hft_10bo_{it} . Using our empirical setup, we evaluate the power of three alternative proxies in explaining the liquidity improvement over the last decade which is attributable to the HFT mostly (Boehmer et al. 2021; Hasbrouck and Saar 2013; Hendershott et al. 2011). We apply an ordinary multivariate regression model which is generally used in the market microstructure literature for determining the impact of market design changes/market variables on market qualities. We use this setup for evaluating the impact of alternative HFT proxies that are defined on the different depth level information of the limit order book.

A primary relationship diagnosis among the variables is made in Table 3. All coefficients in this Pearson's correlation matrix against the null hypothesis $H_0 : \rho = 0$ are significant at 1% level. All three HFT proxies are strongly correlated and inversely related with the spread base liquidity measures but positively correlated with the depth-based measure. We also find that HFT and market fragmentation are positively correlated and market fragmentation has the same relationship with the main response variables as we find for HFT proxies.

Table 3. Pearson’s correlation coefficient matrix. The table reports the Pearson correlation coefficient matrix for the variables calculated on the same balanced sample employed in the regression analyses. The sample comprises 132 stocks for the period December 2005–December 2016 (2624 days). *hft_10bo*, *hft_5bo*, and *hft_bbo* represent the per minute electronic message rate for the 10 best, 5 best, and best bid and offer (BBO) depth levels in the limit order book, respectively, *HHItrd* is the Herfindhal–Hirschman index (HHI) showing the degree of market fragmentation, *spread_bps* is the time-weighted daily quoted spread in basis point, *espread* is the volume weighted effective half-spread in basis point, *depth_bbo* is the average BBO-level quoted depth measured in GBP100, *depth_3bo* is the average cumulative depth up to three best limit prices measured in GBP100, *voltintra* is the intraday mid-price range volatility measured in basis point, *mktcap* is the average market capitalization measured in million GBP, *price* is the daily average price level measured in GBX. All are daily measures constructed from the intraday millisecond records.

Variables	<i>log(hft_10bo)</i>	<i>log(hft_5bo)</i>	<i>log(hft_bbo)</i>	<i>HHItrd</i>	<i>log(spread_bps)</i>	<i>log(espread)</i>	<i>log(depth_bbo)</i>	<i>log(depth_3bo)</i>	<i>Log(mktcap)</i>	<i>log(voltintra)</i>	<i>invprice</i>
<i>log(hft_10bo)</i>	1										
<i>log(hft_5bo)</i>	0.996										
<i>log(hft_bbo)</i>	0.981	0.988									
<i>HHItrd</i>	0.561	0.546	0.503								
<i>log(spread_bps)</i>	−0.762	−0.776	−0.772	−0.402							
<i>Log(espread)</i>	−0.757	−0.766	−0.754	−0.451	0.955						
<i>log(depth_bbo)</i>	0.346	0.381	0.432	−0.095	−0.523	−0.408					
<i>log(depth_3bo)</i>	0.444	0.473	0.513	0.078	−0.593	−0.488	0.969				
<i>Log(mktcap)</i>	0.740	0.754	0.772	0.271	−0.829	−0.773	0.722	0.768			
<i>log(voltintra)</i>	−0.020	−0.027	−0.012	−0.241	0.360	0.372	−0.215	−0.276	−0.234		
<i>invprice</i>	−0.103	−0.108	−0.114	−0.086	0.235	0.257	−0.158	−0.185	−0.156	0.214	1.000

all coefficients are significant at 1% level.

We take three consecutive specifications for defining our target model. We believe that it would help immensely in observing and explaining the impacts of main variables on market liquidity better. We include both stock (firm) fixed effect and time fixed effect (day) to address potential unobserved heterogeneity across firms and time periods in all three specifications. To control the effect of other variables on market liquidity, we also use market capitalization, intraday volatility, and price with proper standardization in the model as they are found to be related with the HFT proxy as the variable of our main interest. The first specification is

$$liq_{it} = \alpha_i + \gamma_t + \beta_1 HFT_{it} + \sigma' X_{it} + \epsilon_{it}, \quad (1)$$

where liq_{it} represents one of the daily (t) market liquidity measures, ($spread_bps_{it}$, $espread_{it}$, $depth_bbo_{it}$, $depth_3bbo_{it}$) for stock i , HFT_{it} represents one of the HFT proxies (hft_bbo_{it} , hft_5bbo_{it} , or hft_10bbo_{it}), the vector X_{it} includes three control variables—log normalized market capitalization ($Log(mktcap)$), log normalized intraday mid price volatility ($Log(voltintra)$), and inverse of daily average prices ($invprice$), which are commonly evident as liquidity determinant in empirical market microstructure literature (Boehmer et al. 2021; Hendershott et al. 2011), α_i is the firm fixed effects, and γ_t is the time-fixed effects.

The expanded second specification, Model 2, includes the daily market fragmentation proxy, $MFrag_{it}$, with HFT_{it} , so that the impact of both HFT and market fragmentation can be assessed in the same model. We use the following specification:

$$Liq_{it} = \alpha_i + \gamma_t + \beta_1 HFT_{it} + \beta_2 MFrag_{it} + \sigma' X_{it} + \epsilon_{it}, \quad (2)$$

where $MFrag_{it}$ is the market fragmentation proxy measured by $HHItrd$.

Our final specification is to include an additional interaction term between HFT and market fragmentation proxies. The final specification is

$$Liq_{it} = \alpha_i + \gamma_t + \beta_1 HFT_{it} + \beta_2 MFrag_{it} + \beta_3 HFT_{it} * MFrag_{it} + \sigma' X_{it} + \epsilon_{it}. \quad (3)$$

The idea of the interaction effect between HFT and market fragmentation arises from the fact that the level of fragmentation and HFT are likely to influence each other (Menkveld 2016).

Our original sample constructs an unbalanced panel of 149 stock. To be on the safe side, and to avoid the probable econometric pitfalls related to the estimation of an unbalanced panel estimation, we construct a balanced panel. The panel is reduced to 132 stocks and 2624 trading days (for the period December 2005–December 2016). All stocks other than italic in Table 1 are included in the balanced panel. Table 4, calculated from the balanced panel, shows the relevant descriptive statistics of regression variables which should be used as the average reference values for all regression estimates. All measures other than 'HHItrd' and 'price' are natural log transformed. We apply the OLS, as an estimation method, using the Newey–West HAC estimator for standard errors, a heteroscedasticity and autocorrelation consistent covariance matrix estimator (lags for autocorrelation are optimally determined).

Table 4. Descriptive Statistics for regression variables. The table presents the descriptive statistics calculated using the balanced sample of 132 stocks for all variables employed in the regression analyses for the period 2005–2016. *hft_10bo*, *hft_5bo*, and *hft_bbo* represent the per minute quote update for the best 10, best 5, and best bid and offer (BBO) depth levels in the limit order, respectively, *HHItrd* is the Herfindhal–Hirschman index (HHI) showing the degree of market fragmentation, *spread_bps* is the time-weighted daily quoted spread in basis point, *espread* is the volume weighted effective half-spread in basis point, *depth_bbo* is the average BBO level quoted depth measured in GBP100, *depth_3bo* is the average cumulative depth up to three best limit price measured in GBP100, *voltintra* is the intraday mid price range volatility measured in basis point, *mktcap* is the average market capitalization measured in million GBP, *price* is the daily average price level measured in GBX. All are daily measures constructed from the intraday millisecond records. The table presents the reference value (mean) for all regression estimates.

Variables	Description	Mean	Median	Std. Dev.	N
<i>spread_bps</i>	quoted half-spreads	18.37	13.14	19.83	346,368
<i>espread</i>	effective half-spread	6.41	4.87	6.92	346,368
<i>depth_bbo</i>	average depth (at BBO/GBP100)	358.15	180.23	869.90	346,368
<i>depth_3bbo</i>	average cumulative depth (best three levels/GBP100)	1636.19	803.05	3613.90	346,368
<i>hft_10bo</i>	electronic message rate per minute (the best 10 depth levels)	114.88	58.75	169.13	346,368
<i>hft_5bo</i>	electronic message rate per minute (the best 5 depth levels)	95.91	51.22	135.38	346,368
<i>hft_bbo</i>	electronic message rate per minute (at BBO)	44.88	25.34	58.11	346,368
<i>HHItrd</i>	Herfindhal Index (proxy for trade fragmenatation)	2.17	2.38	0.74	346,368
<i>mktcap</i>	market capitalization (million GBP)	9666.37	3018.93	17,853.14	346,368
<i>voltintra</i>	Intraday volatility	286.23	216.32	542.32	346,368
<i>price</i>	daily average price (GBX)	934.35	598.07	973.37	346,368

5. Results and Discussions

5.1. HFT Proxy: The 10 Best Prices of the LOB

Tables 5 and 6 report the regression results of the three models where we take the electronic message rate (hft_{it}) constructed on the first 10 levels of the LBO ($hft_{10bo_{it}}$) as the main variable of interest (independent variable) and one of the four liquidity measures ($spread_{bps_{it}}$, $espread_{it}$, $depth_{bbo_{it}}$ and $depth_{3bo_{it}}$) as the response variable. All variables are natural log transformed except *HHItrd* and inverse of the price (*invprice*). All reported coefficients in the Tables 5 and 6 are significant at 1% except for the dependant variable *depth_3bbo*. The coefficients of $\text{Log}(hft_{10bo})$ and *HHItrd* measure the association of HFT intensity and market fragmentation with the liquidity, respectively. We observe from the coefficient of $hft_{10bo_{it}}$ that higher HFT intensity is associated with lower quoted and effective spreads, lower depth (in the both $depth_{bbo_{it}}$ and $depth_{3bo_{it}}$). The coefficients of $M\text{Frag}_{it}$ show that higher fragmentation is associated with higher quoted and effective spreads, lower BBO depth and higher depth at the 3 best prices of the LOB (not significant!). The estimate of -0.278 for $\text{Log}(hft_{10bo_{it}})$ (column I, Table 5) means that, ceteris paribus, 1% increase in the HFT is associated with 0.278% decrease in quoted spread. For instance, a one-standard-deviation increase in HFT from its sample mean of 114.88 messages/minute to 284.01 ($\approx 147.2\%$) would narrow down the quoted spreads by approx. 41% (147.2×0.278), i.e., the sample mean of quoted spreads would fall down from 18.37 bps to 10.85 bps (for descriptive statistics, see Table 4). The coefficient of *HHItrd* is not log transformed, so the estimate of 0.049 (column II) against the log transformed quoted spreads means that, ceteris paribus, a unit increase in *HHItrd*, for example, from the sample mean of 2.17 to 3.17 is associated with 4.9% increase in quoted spreads.

Table 5. Effects of HFT (electronic message traffic rate/minute for the 10 best levels) on market liquidity as measured by relative quoted half-spreads and effective half-spreads). The table reports the panel regression estimates for Models 1–3 where the first two liquidity measures (*spread_bps* and *espread*) are regressed on HFT ($\log(hft_10bo)$), market fragmentation (*HHItrd*) proxy. *hft_10bo* represents the per minute daily quote update for the best 10 depth levels in the limit order book. *HHItrd* is the Herfindhal–Hirschman index (HHI), showing the degree of market fragmentation. The liquidity measures are time-weighted quoted spread (*spread_bps*), volume-weighted effective-half spread (*espread*). Dependent variables, *spread_bps* and *espread* are natural log transformed, all spreads-based measures are in basis point. Control variables are natural log transformed market capitalization ($\text{Log}(mktcap)$), natural log normalized intraday mid-price volatility ($\text{Log}(voltintra)$), and inverse of the average daily price level (*invprice*). The regression is based on a balanced panel of 132 stocks and 2624 days (December 2005–December 2016) and has both time (daily) and stock fixed effects. Coefficient estimates are OLS, t-statistics shown in the parentheses below the coefficient, calculated using Newey–West (HAC) for standard errors, a heteroscedasticity and autocorrelation consistent covariance matrix estimator (lags are optimally determined). *** denotes significance at 1% level.

Explanatory Variables	I	II	III	IV	V	VI
	Model-1	$\text{Log}(\textit{spread_bps})$ Model-2	Model-3	Model-1	$\text{Log}(\textit{espread})$ Model-2	Model-3
<i>log(hft_10bo)</i>	−0.278 *** (−89.68)	−0.279 *** (−90.04)	−0.275 *** (−50.21)	−0.32 *** (−98.95)	−0.321 *** (−99.46)	−0.279 *** (−48.42)
<i>HHItrd</i>		0.049 *** (15.9)	0.057 *** (7.21)		0.061 *** (18.99)	0.139 *** (16.6)
$\log(hft_10bo) * HHItrd$			−0.002 (−1.15)			−0.019 *** (−10.19)
$\log(mktcap)$	−0.214 *** (−52.76)	−0.217 *** (−53.5)	−0.218 *** (−53.34)	−0.141 *** (−32.87)	−0.145 *** (−33.83)	−0.153 *** (−35.53)
$\log(voltintra)$	0.165 *** (56.69)	0.167 *** (58.17)	0.167 *** (57.84)	0.211 *** (62.68)	0.213 *** (64.4)	0.21 *** (62.98)
<i>invprice</i>	17.751 *** (23.43)	17.802 *** (23.41)	17.883 *** (23.38)	20.303 *** (22.68)	20.367 *** (22.69)	21.119 *** (23.35)
stock/firm fixed effect	YES	YES	YES	YES	YES	YES
time fixed effect	YES	YES	YES	YES	YES	YES
Observations	346,368	346,368	346,368	346,368	346,368	346,368
R-Square	0.87	0.87	0.87	0.84	0.84	0.84

We turn now to the interpretation of the interaction term, $\text{Log}(hft_10bo) * HHItrd$, in column III. The interaction term demonstrates the partial effect. In Model 3, the partial effect of hft_10bo_{it} on liquidity depends on the average level of *HHItrd*, and vice versa. To explain, we define here a general expression for the partial effect. The partial effects of HFT_{it} and $MFrag_{it}$ on liq_{it} is defined as $\Delta liq_{it} / \Delta HFT_{it} = \beta_1 + \beta_3 * MFrag_{it}$ and $\Delta liq_{it} / \Delta MFrag_{it} = \beta_2 + \beta_3 * HFT_{it}$, respectively. For interpreting the partial effects, it is necessary that the above two expressions are to be evaluated at some interesting values, generally the mean. The estimated coefficient on the interaction term between HFT and market fragmentation against the response variable $spread_{it}$ is -0.019 (Table 5, Column VI), and the respective estimates for the partial effects of HFT ($\Delta espread / \Delta \text{Log}(hft_10bo)$) and market fragmentation ($\Delta espread / \Delta HHItrd$) are $(-0.279 - 0.019 * 2.17 \approx) -0.32$ and $(0.139 - 0.019 * \text{Log}(114.88) \approx) 0.049$, respectively⁶. These estimates are close to the estimated coefficients on the same variables where the interaction effects are not introduced (column V). We do not report joint significance tests for partial effects, which are trivial as both interaction and main effects are significant. The interaction effect confirms the claim that HFT and market fragmentation are related to each other (Menkveld 2016). The interaction effect implies that some of the possible benefits of HFT intensity on market liquidity are offset by the extra cost of market making that incurs in fragmented markets. Conversely, some extra cost of market fragmentation is also offset by the benefits derived from HFT. It seems that fragmented markets would have more liquidity if there were no HFT.

The impact of *hft_10bo* and *HHItrd* on the average quoted depth (*depth_bbo* and *depth_3bo*) are negative, which imply that both the HFT and fragmentation are associated with less average quoted depth (see Table 6). The positive sign of the coefficient *HHItrd* in

column III implies that more fragmentation is associated with more quoted depth in the deeper level of a limit order book, however, the coefficient for *depth_3bo* is not significant. One might argue that the depleted market liquidity through the quoted depth is likely to outweigh the benefit of the liquidity added through the narrower quoted and effective spreads. We refer to the calibration exercises by [Hendershott et al. \(2011\)](#) to overcome the doubt where the paper concluded that the depth reduction is small relative to the narrowing of the spread.

The coefficients of the control variables, in both Tables 5 and 6, have the expected signs and are significant at the 1% level. The large market capitalized stocks are associated with lower quoted and effective spreads, lower price impacts, greater depth, and higher realized spreads. The inverse price coefficient implies that a stock with higher price is associated with lower quoted and effective spreads, higher depth and lower realized spreads and price impacts. The positive estimate of the volatility coefficient implies that a higher intraday volatility increases quoted and effective spreads, provides greater depth at BBO, also associated with higher price impact and lower realized spread.

Table 6. Effect of HFT (electronic message rate/minute for the 10 best levels) on market liquidity as measured by the market depth at BBO and the cumulative market depth for 3 best depth levels. The table reports the panel regression results of Models 1–3 where two depth-based liquidity measures (*depth_bbo* and *depth_3bo*) are regressed on HFT (*hft_10bo*) and market fragmentation (*HHItrd*) proxy. *hft_10bo* represents the per minute daily quote update in the 10 best depth levels of the limit order book. *HHItrd* is the Herfindhal–Hirschman index (HHI), shows the degree of market fragmentation. The liquidity measures (response variables) are the average quoted depth at best limit price (*depth_bbo*), and the accumulated average quoted depth up to the best three limit price (*depth_3bo*). All explanatory variables are natural log transformed, and depth measures are in GBP100. Control variables are log market capitalization (*Log(mktcap)*), natural log normalized intraday mid-price volatility (*Log(volintra)*), and inverse of the average daily price level (*invprice*). The regression is based on a balanced panel of 132 stocks and 2624 days (December 2005–December 2016) and has both time (day) and stock (firm) fixed effects. Coefficient estimates are OLS, t-statistics shown in the parentheses below the coefficient, calculated using Newey–West (HAC) for standard errors, a heteroscedasticity and autocorrelation consistent covariance matrix estimator (lags are optimally determined). *** denotes significance at 1% level.

Explanatory Variables	I	II	III	IV	V	VI
	Model-1	<i>Log(depth_bbo)</i> Model-2	Model-3	Model-1	<i>Log(depth_3bo)</i> Model-2	Model-3
<i>log(hft_10bo)</i>	−0.284 *** (−45.35)	−0.282 *** (−45.25)	−0.183 *** (−16.13)	−0.295 *** (−43.27)	−0.295 *** (−43.22)	−0.285 *** (−22.7)
<i>HHItrd</i>		−0.107 *** (−19.94)	0.078 *** (4.76)		0.001 (0.21)	0.019 (1.05)
<i>Log(hft_10bo) * HHItrd</i>			−0.045 *** (−10.99)			−0.004 (−0.98)
<i>log(mktcap)</i>	0.816 *** (87.13)	0.823 *** (87.53)	0.804 *** (87.28)	0.864 *** (83.64)	0.864 *** (83.4)	0.862 *** (84.29)
<i>log(volintra)</i>	0.041 *** (10.77)	0.037 *** (9.52)	0.029 *** (7.39)	0.003 (0.73)	0.003 (0.74)	0.002 (0.55)
<i>invprice</i>	1.067 (0.56)	0.957 (0.51)	2.739 (1.46)	−0.144 (−0.08)	−0.143 (−0.08)	0.032 (0.02)
stock/firm fixed effect	YES	YES	YES	YES	YES	YES
time fixed effect	YES	YES	YES	YES	YES	YES
observations	346,368	346,368	346,368	346,368	346,368	346,368
R-Square	0.82	0.82	0.83	0.81	0.81	0.81

5.2. HFT Proxy: The 5 Best Prices of the LOB

Tables 7 and 8 report the regression results of the HFT proxy defined on the five best prices of the limit order book. All the coefficient estimates in Table 7 for three models are significant and higher than those estimated for the proxy *hft_10bo* in Table 5. Interestingly, we see that HFT proxy based on the 5 LOB best prices is showing a stronger impact on market liquidity than the one we defined on the 10 best prices. On the contrary, the regression coefficient for all models run on depth-level liquidities (Table 8), are weaker than what we find in Table 6. The estimated coefficient for the market fragmentation and the interaction between HFT and market fragmentation beyond the BBO are not found to be significant, but that for the HFT are significant and stronger for the depth beyond the BBO. Our evidence agrees with the evidence provided in AMF (2017).

Table 7. Effect of HFT (electronic message traffic rate/minute for 5 best levels) on market liquidity as measured by quoted spreads and effective spreads). The table presents the panel regression results of Models 1–3 where first two liquidity measures (*spread_bps* and *espread*) are regressed on HFT ($\log(hft_5bo)$) and market fragmentation (*HHItrd*) proxy. *hft_5bo* represents the per minute daily quote update for the best 5 depth levels in the limit order book. *HHItrd* is the Herfindhal–Hirschman index (HHI), showing the degree of market fragmentation. The liquidity measures are time-weighted quoted spread (*spread_bps*), volume-weighted effective-half spread (*espread*). Dependent variables, *spread_bps*, *espread* are natural log transformed, all spreads-based measures are in basis point. Control variables are lnatural og transformed market capitalization ($\log(mktcap)$), natural log normalized intraday mid-price volatility ($\log(voltintra)$) and inverse of the average daily price level (*invprice*). The regression is based on a balanced panel of 132 stocks and 2624 days (December 2005–December 2016) and has both time (daily) and stock fixed effects. Coefficient estimates are OLS, t-statistics shown in the parentheses below the coefficient, calculated using Newey–West (HAC) for standard errors, a heteroscedasticity and autocorrelation consistent covariance matrix estimator (lags are optimally determined). *** denotes significance at 1% level.

Explanatory Variables	I	II	III	IV	V	VI
	Model-1	Log(<i>spread_bps</i>)		Model-1	Log(<i>espread</i>)	
		Model-2	Model-3		Model-2	Model-3
<i>log(hft_5bo)</i>	−0.288 *** (−92.96)	−0.289 *** (−93.21)	−0.283 *** (−54.74)	−0.322 *** (−99.98)	−0.324 *** (−100.38)	−0.282 *** (−51.56)
<i>HHItrd</i>		0.052 *** (17.72)	0.065 *** (9.09)		0.065 *** (20.72)	0.144 *** (18.81)
<i>log(hft_5bo) * HHItrd</i>			−0.003 * (−1.96)			−0.02 *** (−11.24)
<i>log(mktcap)</i>	−0.198 *** (−49.14)	−0.201 *** (−49.93)	−0.202 *** (−49.97)	−0.127 *** (−29.43)	−0.131 *** (−30.41)	−0.139 *** (−32.17)
<i>log(voltintra)</i>	0.167 *** (58.38)	0.169 *** (59.99)	0.168 *** (59.69)	0.21 *** (63.96)	0.213 *** (65.83)	0.209 *** (64.35)
<i>invprice</i>	17.159 *** (22.65)	17.205 *** (22.68)	17.335 *** (22.7)	19.847 *** (21.99)	19.904 *** (22.04)	20.694 *** (22.83)
stock/firm fixed effect	YES	YES	YES	YES	YES	YES
time fixed effect	YES	YES	YES	YES	YES	YES
Observations	346,368	346,368	346,368	346,368	346,368	346,368
R-Square	0.87	0.87	0.87	0.84	0.84	0.84

Table 8. Effect of HFT (electronic message traffic rate/minute for 5 best levels) on market liquidity as measured by the marker depth at BBO and the cumulative market depth for 3 best depth levels). The table presents the panel regression results of Models 1–3 where two depth-based liquidity measures (*depth_bbo* and *depth_3bo*) are regressed on HFT (*hft_5bo*) and market fragmentation (*HHItrd*) proxy. *hft_5bo* represents the per minute quote update for the best 5 depth levels in the limit order book. *HHItrd* is the Herfindhal–Hirschman index (HHI), showing the degree of market fragmentation. The liquidity measures are the average quoted depth at best limit price (*depth_bbo*), the accumulated average quoted depth up to the best three limit price (*depth_3bo*). All dependent variables are log transformed and depth measures are in 100GBP. Control variables are log market capitalization (*Log(mktcap)*), log normalized intraday mid-price volatility (*Log(voltintra)*) and inverse of the average daily price level (*invprice*). The regression is based on a balanced panel of 132 stocks and 2624 days (December 2005–December 2016) and has both time (daily) and stock fixed effects. Coefficient estimates are OLS, t-statistics shown in the parentheses below the coefficient, calculated using Newey–West (HAC) for standard errors, a heteroscedasticity and autocorrelation consistent covariance matrix estimator (lags are optimally determined). *** and * denote significance at 1% and 10% levels, respectively.

	I	II	III	IV	V	VI
Explanatory Variables	Model-1	Log(<i>depth_3bo</i>) Model-2	Model-3	Model-1	Log(<i>depth_bbo</i>) Model-2	Model-3
<i>log(hft_5bo)</i>	−0.25 *** (−40.36)	−0.247 *** (−40.14)	−0.155 *** (−14.76)	−0.268 *** (−39.84)	−0.268 *** (−39.78)	−0.264 *** (−22.66)
<i>HHItrd</i>		−0.105 *** (−19.51)	0.068 *** (4.5)		0.003 (0.52)	0.01 (0.62)
<i>log(hft_5bo) * HHItrd</i>			−0.044 *** (−11.27)			−0.002 (−0.43)
<i>log(mktcap)</i>	0.811 *** (85.82)	0.818 *** (86.18)	0.801 *** (86.22)	0.863 *** (82.97)	0.863 *** (82.75)	0.862 *** (83.75)
<i>log(voltintra)</i>	0.029 *** (7.69)	0.025 *** (6.42)	0.018 *** (4.59)	−0.007 * (−1.71)	−0.007 * (−1.68)	−0.007 * (−1.75)
<i>invprice</i>	1.584 (0.82)	1.492 (0.77)	3.227 * (1.69)	0.173 (0.09)	0.176 (0.09)	0.249 (0.13)
stock/firm fixed effect	YES	YES	YES	YES	YES	YES
time fixed effect	YES	YES	YES	YES	YES	YES
Observations	346,368	346,368	346,368	346,368	346,368	346,368
R-Square	0.82	0.82	0.82	0.81	0.81	0.81

5.3. HFT Proxy: The First Best Price (BBO) of the LOB

For the last HFT proxy, *hft_bbo*, Tables 9 and 10 report all the estimated regression coefficients. All estimated coefficients for the first two spread base liquidity measures (Table 9) are significant (except one in column III) and show the weakest association with the HFT and market fragmentation among all the estimates that we estimated so far in Tables 5–10. The impact of market fragmentation on the depth base liquidity is not established if we rely on the HFT proxy at the BBO-level information. The results go against the evidence we find from one of the important studies in the European market on fragmentation by Gresse (2017). The results we obtained by employing the HFT proxy defined on the BBO level information has turned out to be the weakest and most fragile.

5.4. Which Depth Levels of LOB Should We Rely on?

At this point, we would like to narrow down our choices on three specifications. The analyses we have made so far make it clear that the third specification (Model 3) provides the most reliable estimate. We would also like to single out a market liquidity measure so that the results can be concluded more significantly. Here, our choice is the effective spreads due to the following reasons. Firstly, the results demonstrated so far are similar for both quoted spreads and effective spreads but more robust for the latter one as a dependent variable. Secondly, the effective spread is considered a more sound measure of liquidity theoretically. We find that on average effective spreads are lower than the quoted spreads (see Table 4). For understanding the causes for lower effective spreads,

the remark of the Petersen and Fialkowski (1994) is significant: “when trades are executed inside the posted bid-ask spread, the posted spread is no longer an accurate measure of transaction costs faced by investor”. The same study also shows that the effective spread averages half the posted spread. This phenomenon can be generalized and persists in every market including emerging ones (Ahn et al. 2018). Hagströmer (2021) confirms the diverse application of effective spreads as a measure of trading cost, and a benchmark measure of many regulators. We summarize next the regression results for all HFT proxies regressed on effective spreads.

Table 11 summarizes the results of Model 3 where liquidity (effective spreads) is regressed on alternative HFT proxies—the electronic message rate defined on the three different levels of LOB information. It is clear that the highest impact of HFT, market fragmentation, and their interaction on the liquidity is given for the electronic message rate calculated on the 5 best prices of LOB (*hft_5bo*). It appears that neither *hft_10bo* (the electronic message rate calculated on the 10 best prices) nor *hft_bbo* (the electronic message rate calculated on the first best prices) provides as many variation in regressions for explaining the variations in the liquidity measures as supplied by *hft_5bo* (the electronic message rate calculated on the five best prices). The evidence supports that HFT participates beyond the BBO as reported in AMF (2017), and confirms that HFTs provide both tight (marketable) and wider (non-marketable) quotes. Our study also extends the evidence provided in AMF (2017) to show that HFTs’ participation at the 10 best prices, i.e., at too far from the BBO, is not significant.

Table 9. Effect of HFT (electronic message traffic rate/minute at BBO) on market liquidity as measured by relative quoted spreads and effective half-spreads. The table presents the panel regression results of Models 1–3 where the first two liquidity measures (*spread_bps* and *espread*) are regressed on HFT ($\log(hft_bbo)$), market fragmentation (*HHItrd*) proxy. *hft_bbo* represents the per minute quote update for the BBO depth levels in the limit order book. *HHItrd* is the Herfindhal–Hirschman index (HHI), shows the degree of market fragmentation. The liquidity measures are time-weighted quoted spread (*spread_bps*), volume-weighted effective-half spread (*espread*). Dependent variables, *spread_bps*, *espread* are log transformed, all spreads-based measures are in basis point. Control variables are natural log transformed market capitalization ($\log(mktcap)$), natural log normalized intraday mid-price volatility ($\log(voltintra)$), and inverse of the average daily price level (*invprice*). The regression is based on a balanced panel of 132 stocks and 2624 days (December 2005–December 2016) and has both time (daily) and stock fixed effects. Coefficient estimates are OLS, t-statistics shown in the parentheses below the coefficient, calculated using Newey–West (HAC) for standard errors, a heteroscedasticity and autocorrelation consistent covariance matrix estimator (lags are optimally determined). *** denotes significance at 1% level.

Explanatory Variables	I		II	III	IV	V		VI
	<i>Log(spread_bps)</i>					<i>Log(espread)</i>		
	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-3	Model-3
<i>log(hft_bbo)</i>	−0.269 *** (−81.77)	−0.269 *** (−81.87)	−0.265 *** (−48.72)	−0.291 *** (−84.2)	−0.292 *** (−84.41)	−0.254 *** (−42.99)		
<i>HHItrd</i>		0.047 *** (15.35)	0.054 *** (8.35)		0.059 *** (17.78)	0.118 *** (17.05)		
<i>log(hft_bbo) * HHItrd</i>			−0.002 (−1.08)			−0.018 *** (−9.68)		
<i>log(mktcap)</i>	−0.203 *** (−47.98)	−0.207 *** (−48.64)	−0.207 *** (−48.46)	−0.138 *** (−29.98)	−0.142 *** (−30.81)	−0.148 *** (−32.15)		
<i>log(voltintra)</i>	0.162 *** (55.6)	0.164 *** (57)	0.164 *** (56.79)	0.202 *** (60.4)	0.204 *** (62.03)	0.202 *** (60.66)		
<i>invprice</i>	17.29 *** (20.89)	17.338 *** (20.87)	17.406 *** (20.85)	20.246 *** (20.3)	20.306 *** (20.3)	20.963 *** (20.93)		
stock/firm fixed effect	YES	YES	YES	YES	YES	YES		YES
time fixed effect	YES	YES	YES	YES	YES	YES		YES
Observations	346,368	346,368	346,368	346,368	346,368	346,368		346,368
R-Square	0.87	0.87	0.87	0.83	0.83	0.83		0.83

Table 10. Effect of HFT (electronic message traffic rate/minute at BBO) on market liquidity as measured by the marker depth at BBO and the cumulative market depth for 3 best depth levels). The table presents the panel regression results of Models 1–3 where two depth-based liquidity measures (*depth_bbo* and *depth_3bo*) are regressed on HFT (*hft_bbo*) and market fragmentation (*HHItrd*) proxy. *hft_bbo* represents the per minute daily quote update for the BBO in the limit order book. *HHItrd* is the Herfindhal–Hirschman index (HHI), which shows the degree of market fragmentation. The liquidity measures are the average quoted depth at best limit price (*depth_bbo*), the accumulated average quoted depth up to the best three limit price (*depth_3bo*). All dependent variables are natural log transformed and depth measures are in 100GBP. Control variables are natural log market capitalization (*Log(mktcap)*), natural log normalized intraday mid-price volatility (*Log(voltintra)*), and inverse of the average daily price level (*invprice*). The regression is based on a balanced panel of 132 stocks and 2624 days (December 2005–December 2016) and has both time (daily) and stock fixed effects. Coefficient estimates are OLS, t-statistics shown in the parentheses below the coefficient, calculated using Newey–West (HAC) for standard errors, a heteroscedasticity and autocorrelation consistent covariance matrix estimator (lags are optimally determined). ***, **, * denote significance at 1%, 5%, and 10% levels, respectively.

	I	II	III	IV	V	VI
Explanatory Variables	Model-1	Log(<i>depth_bbo</i>) Model-2	Model-3	Model-1	Log(<i>depth_3bo</i>) Model-2	Model-3
<i>log(hft_bbo)</i>	−0.148 *** (−24.55)	−0.146 *** (−24.45)	−0.07 *** (−6.76)	−0.184 *** (−27.66)	−0.184 *** (−27.63)	−0.188 *** (−16.19)
<i>HHItrd</i>		−0.112 *** (−20.37)	0.007 (0.52)		−0.004 (−0.57)	−0.01 (−0.74)
<i>log(hft_bbo) * HHItrd</i>			−0.037 *** (−9.4)			0.002 (−0.5)
<i>log(mktcap)</i>	0.764 *** (79.74)	0.771 *** (80.22)	0.758 *** (80.13)	0.825 *** (−78.28)	0.825 *** (−78.11)	0.826 *** (−78.92)
<i>log(voltintra)</i>	−0.002 (−0.46)	−0.007 * (−1.72)	−0.012 *** (−3.05)	−0.032 *** (−7.89)	−0.032 *** (−7.93)	−0.032 *** (−7.85)
<i>invprice</i>	3.97 * (1.92)	3.856 * (1.88)	5.173 ** (2.53)	2.066 (−1.02)	2.063 (−1.02)	1.986 (−0.97)
stock/firm fixed effect	YES	YES	YES	YES	YES	YES
time fixed effect	YES	YES	YES	YES	YES	YES
Observations	346,368	346,368	346,368	346,368	346,368	346,368
R-Square	0.82	0.82	0.82	0.81	0.81	0.81

We would like to explain our results more rigorously. The evidence we provide for the European market in the Section 4.2.3 along with recent studies in US (e.g., Hasbrouck and Saar 2013; Hendershott et al. 2011) and international markets (e.g., Boehmer et al. 2021) show a sharp decline in transaction costs over the past. The same studies have attributed this decline of transaction cost mostly to the simultaneous rise of high-frequency traders at the global market places. In the absence of directly flagged and reliable high-frequency data, recent studies have deliberately used the unflagged LOB data for measuring HFT proxies. We find it worth mentioning again that case studies specific to the HFT participation at different level of market depths (e.g., AMF 2017; Hendershott and Riordan 2013) using exchange supplied flagged data document that HFTs use both marketable (HFTs’ share is 52% to 70.8%) and nonmarketable/passive (HFTs’ share is 64% to 79.3%) quotes and trades in the LOB’s participation. A marketable quote is one which finds the counter party and gets executed immediately. The first level of market depth or the BBO represents the marketable quotes in a LOB. A passive or nonmarketable quote represents all the orders queued behind the BBO. We have already discussed (see Section 4.2.4) the strength of the electronic message rate (*hft_{it}*) over other measures as a candidate of HFT proxy. Now, the issue which remains to be settled is the level of limit order books information that should be used for constructing HFT proxies which can exhaust HFTs’ participation in the LOB mostly. Otherwise, it poses either the risk of missing important HFT footprints which makes the proxy biased or underidentified, or imposes the burden of handling

voluminousness data that eventually impact both financial and non-financial resource commitments for a study.

Prior HFT studies using proprietary LOB data predominantly have used the first level of market depth data for constructing their proxies. Based on the evidence we provide from the related studies, it manifests that defining the electronic message rate only on the BBO level information does not depict a full HFT image, and raises the concerns of under or over identification. As we reported earlier, among the proxies defined on three chosen levels of market depth, *hft_bbo* shows the weakest association with the HFT and market fragmentation (see Tables 5–10) that is consistent with our concern of under-identification problem. If our conjecture is valid, it implies that *hft_bbo* underestimates the coefficient for the liquidity variable (effective spreads) consistently, and the amount of underestimation is on average about 10%⁷ and 9% compare to the same for *hft_5bo* and *hft_10bo*, respectively.

Why is the use of BBO level data so widespread in HFT research? There are two aspects we need to address in finding the answer to this question. Firstly, the supply of HFT data is quite limited with regards to both time and instrument coverages. Secondly, data management remains a challenge due to its voluminous nature. The more the data fields get involved in a dataset, the higher the data management burden, since most of the HFT data are recorded at the granularity of millisecond or microsecond. Generally, a small file involves millions of records. As a result, the use of LOB data beyond the BBO is not seen frequently in HFT research. After the seminal study of [Hendershott et al. \(2011\)](#), many papers have followed in the same path, and there is reasonable ground to believe that the limitations associated with the HFT data have delayed this valid question that there should be a LOB level up to whose use the proxy can become a valid measure of HFT.

Why is the use of 10 best level data counterintuitive? The arguments and evidence we have provided so far may give the impression that the more the information from the limit order book we incorporate the better the HFT measure becomes, but this is not correct. HFTs participate in nonmarketable quotes which are essentially queued behind the BBO. Now, the fundamental question is—how passive are the quotes? Since HFTs are fast, they can minimize the adverse selection costs by their market monitoring technology. For this reason, expecting a nonmarketable quote from HFT which is too passive is not intuitive. Again, the handling of 10 level market depth data is quite costly in terms of the use of computing resources and the data subscription costs. The estimates we find from the empirical market microstructure models (see Tables 5–10) for the liquidity variables against the proxy *hft_10bo* are weaker than those we get for the *hft_5bo*. The results support our conjecture that the use of too much information for the construction of HFT proxy increases the implied costs without any benefits, that we refer to as welfare damaging.

The evidence we provide here has a profound impact on the selection of the LOB best prices for constructing HFT proxies. Our results indicate that the electronic message rate (*hft_{it}*) defined by the five best prices provides a better estimate than that only defined by the BBO prices. It also suggests that taking the best prices after a certain level of the LBO (e.g., five best prices) for defining the electronic message rate (*hft_{it}*) is counterintuitive. We do not claim that the electronic message rate (*hft_{it}*) defined on the five best prices of the limit order book is optimal. The methodology we adopt in this paper is like an exercise where we take three choices from an order book—two edges of the LOB (*hft_bbo* and *hft_10bo*) and a mid-point (*hft_5bo*), to show how the regression estimates of the generally established market microstructure models change with the different uses of LOB data. We could have expanded our choices on more LOB points in doing this calibration, but our idea in this paper is to raise the issue and put it forward as an agenda for future research. We admit that the scope of this paper does not allow us to address this issue fully in one paper and it is also true that we do not have the supply of both flagged and unflagged HFT data by using which we could have come up with a definitive answer.

Table 11. Effects of HFT on liquidity: summary estimates. The table summarize the estimates reported in Tables 5–10 for Model 3. *** and * denote significance at 1%, and 10% levels, respectively.

LOB Depth Level	Log(<i>spread_bps</i>)			Log(<i>espread</i>)			Log(<i>depth_bbo</i>)			Log(<i>depth_3bo</i>)		
	<i>hft_bbo</i>	<i>hft_5bo</i>	<i>hft_10bo</i>	<i>hft_bbo</i>	<i>hft_5bo</i>	<i>hft_10bo</i>	<i>hft_bbo</i>	<i>hft_5bo</i>	<i>hft_10bo</i>	<i>hft_bbo</i>	<i>hft_5bo</i>	<i>hft_10bo</i>
<i>log(HFT_{it})</i>	−0.265 *** (−48.72)	−0.283 *** (−54.74)	−0.275 *** (−50.21)	−0.254 *** (−42.99)	−0.282 *** (−51.56)	−0.279 *** (−48.42)	−0.07 *** (−6.76)	−0.155 *** (−14.76)	−0.183 *** (−16.13)	−0.188 *** (−16.19)	−0.264 *** (−22.66)	−0.285 *** (−22.7)
<i>HHItrd</i>	0.054 *** (8.35)	0.065 *** (9.09)	0.057 *** (7.21)	0.118 *** (17.05)	0.144 *** (18.81)	0.139 *** (16.6)	0.007 (0.52)	0.068 *** (4.5)	0.078 *** (4.76)	−0.01 (−0.74)	0.01 (0.62)	0.019 (1.05)
<i>log(HFT_{it}) * HHItrd</i>	−0.002 (−1.08)	−0.003 * (−1.96)	−0.002 (−1.15)	−0.018 *** (−9.68)	−0.02 *** (−11.24)	−0.019 *** (−10.19)	−0.037 *** (−9.4)	−0.044 *** (−11.27)	−0.045 *** (−10.99)	0.002 (−0.5)	−0.002 (−0.43)	−0.004 (−0.98)
<i>log(mktcap)</i>	−0.207 *** (−48.46)	−0.202 *** (−49.97)	−0.218 *** (−53.34)	−0.148 *** (−32.15)	−0.139 *** (−32.17)	−0.153 *** (−35.53)	0.758 *** (80.13)	0.801 *** (86.22)	0.804 *** (87.28)	0.826 *** (−78.92)	0.862 *** (83.75)	0.862 *** (84.29)
<i>log(voltintra)</i>	0.164 *** (56.79)	0.168 *** (59.69)	0.167 *** (57.84)	0.202 *** (60.66)	0.209 *** (64.35)	0.21 *** (62.98)	−0.012 *** (−3.05)	0.018 *** (4.59)	0.029 *** (7.39)	−0.032 *** (−7.85)	−0.007 * (−1.75)	0.002 (0.55)
<i>invprice</i>	17.406 *** (20.85)	17.335 *** (22.7)	17.883 *** (23.38)	20.963 *** (20.93)	20.694 *** (22.83)	21.119 *** (23.35)	5.173 ** (2.53)	3.227 * (1.69)	2.739 (1.46)	1.986 (−0.97)	0.249 (0.13)	0.032 (0.02)
stock/firm fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
time fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
observations	346,368	346,368	346,368	346,368	346,368	346,368	346,368	346,368	346,368	346,368	346,368	346,368
R-Square	0.87	0.87	0.87	0.83	0.84	0.84	0.82	0.82	0.83	0.81	0.81	0.81

6. Conclusions

HFT research has been trending in the last decade. In the absence of a uniform HFT identification strategy, we see that the evidence on empirical HFT studies is somewhat conflicting and inconclusive. In this paper, we attempt to evaluate HFT proxies that have been employed in many prior studies. The aim is to check with evidence whether they all have the same or different HFT identifying qualities. Our study has benefited from millisecond time-stamped rich datasets supplied by TRTH with data for 149 LSE-listed large market capitalized stocks for the period of 2005 to 2016. The datasets enable us to construct many HFT proxies with alternative definitions, different measures for liquidity, and a proxy for market fragmentation.

Our study confirms that the post-MiFID era of the European equity market, particularly LSE stocks, has documented intense HFT activities along with high fragmentation in trading volume across markets. The joint forces of HFT and market fragmentation have made a significant impact on the improvement of liquidity in the European market. However, gauging the HFT through proxies remains an issue here.

Our analyses show that some HFT proxies such as *algo_trad*, used in [Hendershott et al. \(2011\)](#), and order-to-trade ratio (*order_to_trade*) may not be used invariably in all market settings, and their interpretations also require proper attention. It reveals that their cross-sectional and time-series comparisons do not stand on the same mathematical arguments and are somewhat dubious. The existing studies, namely, [Bouveret et al. \(2014\)](#), have also raised concern over using the order-to-trade ratio as an HFT proxy. Apart from that, the construction of some other HFT proxies such as, 'life time of order' [Bouveret et al. \(2014\)](#), 'overnight inventory position' ([Kirilenko et al. 2017](#)), or 'strategic runs' ([Hasbrouck and Saar 2013](#)), are normally not possible for the limited availability of data fields in an HFT dataset.

Popular HFT proxies generally constructed from the trade and quote data are also not uniformly defined but mostly rely on the first best prices of the limit order book. For example, [Boehmer et al. \(2021\)](#); [Hendershott et al. \(2011\)](#); [Upson and Van Ness \(2017\)](#); and [Conrad et al. \(2015\)](#) have used only the BBO-level information for defining their proxies, whereas our results along with the evidence provided in the studies like [AMF \(2017\)](#) support that HFTs participate beyond the BBO and provide both tight (marketable) and wider (non-marketable) quotes. The results also show that the electronic message rate defined on the BBO level information provides the weakest estimate among the three alternative depth levels used in this study. Our results also suggest that taking the best prices after a certain level (e.g., 5 best prices) of the LBO for defining the HFT proxy is counterintuitive. We find that the electronic message rate (hft_{it}) defined on the 5 best prices of the limit order book provides the largest variations in regressions for explaining the European market quality changes for the period 2005–2016.

The evidence we document in this paper supports the idea that HFT provides liquidity at a deeper level than the best bid and offer price in the limit order book. However, HFTs do not participate in quoting at the very deep of a limit order book (e.g., 10 levels of market depth). The HFT proxy defined on the best five level of limit order book updates explains the market liquidity changes most.

HFTs are diverse in their use of trading strategies ([Biais and Foucault 2014](#); [Hagströmer and Nordén 2013](#)). The HFT proxies we finally sorted in this paper are likely to capture the relative dominance of a subset of HFT which we believe to be 'market making'. The evidence we provide here has important implications for the empirical HFT study which are not supported by the HFT-flagged datasets. We do not claim that the electronic message rate (hft_{it}) defined on the 5 best prices of the limit order book is optimal, since we take here only two extremes and one median point of the limit order book information for performing the analyses. We believe that any similar study designed on additional informative points of the limit order book may come out with a more precise recommendations. Our paper provides a general guideline in developing HFT proxies such as the electronic message rate (hft_{it}) where one should not rely on the BBO level information or extract too much

information from the limit order book owing to the fact that every additional piece of information essentially requires more commitment of time and financial resources.

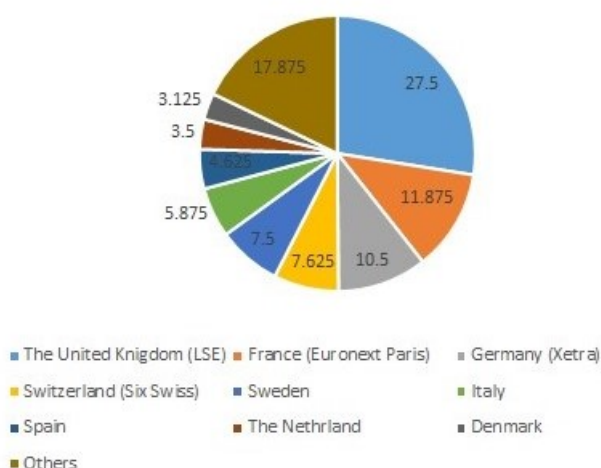
Funding: The idea for this paper and some of the analyses draw upon Chapter 2 of my Ph.D. thesis at University of Naples Federico II. I thank Marco Pagano for insightful comments and suggestions. Financial support from the University of Naples Federico II is gratefully acknowledged.

Data Availability Statement: Restrictions apply to the availability of these data. Raw data were obtained from Thomson Reuters Tick History (TRH) owned by Securities Industry Research Centre of Asia-Pacific (SIRCA) through the European Capital Markets Cooperative Research Centre (ECMCRC), and are available from the authors with the permission of European Capital Markets Cooperative Research Centre and SIRCA.

Conflicts of Interest: The author declares no conflict of interest.

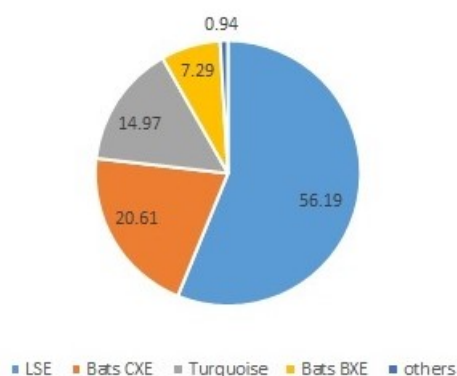
Appendix A

STOXX 800 Compositions: Primary Listing Venues



(a)

Lit Trading Market Share: LSE Stocks



(b)

Figure A1. Sample universe: STOXX 800. (a) STOXX 800 composition. (b) Lit Trading Market share (Source: <https://fragmentation.fidessa.com/>, accessed on 15 January 2018).

Table A1. Snapshot of LOB Updates: 10 level market depth. The table lists 10 consecutive updates of market depth for the security “BATS.L” in the limit order book. This specimen snapshot record is taken from the market depth file supplied from TRTH. The market depth is listed upto its best 10 level. Every update shows the change in one or more of the fields listed here.

#RIC Date[L] Time[L] Type	BATS.L 29-Sep-16 4:11:45 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth	BATS.L 29-Sep-16 4:11:46 PM Market Depth
L1-BidPrice	4941	4941	4941	4941	4941	4941	4941	4941	4941	4941
L1-BidSize	1044	1044	897	897	897	697	697	697	697	547
L1-BuyNo	6	6	5	5	5	4	4	4	4	3
L1-AskPrice	4942	4942	4942	4941.5	4941.5	4941.5	4941.5	4941.5	4941.5	4941.5
L1-AskSize	1083	1083	200	200	200	200	200	299	434	434
L1-SellNo	7	7	7	7	7	8	8	8	8	7
L2-BidPrice	4940.5	4940.5	4940.5	4940.5	4940.5	4940.5	4940.5	4940.5	4940.5	4940.5
L2-BidSize	2098	2098	2098	2098	2098	2098	2098	2098	2098	2098
L2-BuyNo	14	14	14	14	14	14	14	14	14	14
L2-AskPrice	4942.5	4942.5	4942.5	4942	4942	4942	4942	4942	4942	4942
L2-AskSize	1769	1769	1083	1083	1083	1103	1103	1103	1103	1103
L2-SellNo	13	13	13	13	13	8	8	8	8	8
L3-BidPrice	4940	4940	4940	4940	4940	4940	4940	4940	4940	4940
L3-BidSize	1938	1938	1938	1938	1938	1938	1938	1938	1938	1938
L3-BuyNo	13	13	13	13	13	13	13	13	13	13
L3-AskPrice	4943	4943	4943	4942.5	4942.5	4942.5	4942.5	4942.5	4942.5	4942.5
L3-AskSize	2048	2048	2048	1769	1769	1769	1769	1769	1769	1769
L3-SellNo	13	13	13	13	13	13	13	13	13	13
L4-BidPrice	4939.5	4939.5	4939.5	4939.5	4939.5	4939.5	4939.5	4939.5	4939.5	4939.5
L4-BidSize	2415	2415	2415	2415	2415	2415	2415	2415	2415	2415
L4-BuyNo	14	14	14	14	14	14	14	14	14	14
L4-AskPrice	4943.5	4943.5	4943.5	4943	4943	4943	4943	4943	4943	4943
L4-AskSize	2007	2007	2007	2048	2048	2048	2048	2048	2048	2048
L4-SellNo	13	13	13	13	13	13	13	13	13	13
L5-BidPrice	4939	4939	4939	4939	4939	4939	4939	4939	4939	4939
L5-BidSize	2310	2310	2310	2310	2310	2310	2310	2310	2310	2310
L5-BuyNo	12	12	12	12	12	12	12	12	12	12
L5-AskPrice	4944	4944	4944	4943.5	4943.5	4943.5	4943.5	4943.5	4943.5	4943.5
L5-AskSize	2192	2692	2692	2007	2007	2007	2007	2007	2007	2007
L5-SellNo	13	14	14	13	13	13	13	13	13	13
L6-BidPrice	4938.5	4938.5	4938.5	4938.5	4938.5	4938.5	4938.5	4938.5	4938.5	4938.5
L6-BidSize	2140	2140	2140	2140	2140	2140	2140	2140	2140	2140
L6-BuyNo	9	9	9	9	9	9	9	9	9	9
L6-AskPrice	4944.5	4944.5	4944.5	4944	4944	4944	4944	4944	4944	4944
L6-AskSize	1681	1681	1681	2692	2692	2692	2692	2692	2692	2692
L6-SellNo	11	11	11	14	14	14	14	14	14	14
L7-BidPrice	4938	4938	4938	4938	4938	4938	4938	4938	4938	4938
L7-BidSize	3044	3044	3044	3044	3044	3044	3044	3044	3044	3044
L7-BuyNo	9	9	9	9	9	9	9	9	9	9
L7-AskPrice	4945	4945	4945	4944.5	4944.5	4944.5	4944.5	4944.5	4944.5	4944.5
L7-AskSize	3363	3363	3363	1681	1681	1681	1681	1681	1681	1681
L7-SellNo	13	13	13	11	11	11	11	11	11	11
L8-BidPrice	4937.5	4937.5	4937.5	4937.5	4937.5	4937.5	4937.5	4937.5	4937.5	4937.5
L8-BidSize	923	923	923	923	923	923	923	923	923	923
L8-BuyNo	5	5	5	5	5	5	5	5	5	5
L8-AskPrice	4945.5	4945.5	4945.5	4945	4945	4945	4945	4945	4945	4945
L8-AskSize	1881	1881	1881	3363	3363	3363	3363	3363	3363	3363
L8-SellNo	11	11	11	13	13	13	13	13	13	13
L9-BidPrice	4937	4937	4937	4937	4937	4937	4937	4937	4937	4937
L9-BidSize	1109	1109	1109	1109	1109	1109	1109	1109	1109	1109
L9-BuyNo	4	4	4	4	4	4	4	4	4	4
L9-AskPrice	4946	4946	4946	4945.5	4945.5	4945.5	4945.5	4945.5	4945.5	4945.5
L9-AskSize	1071	1071	1071	1881	1681	1681	1681	1681	1681	1681
L9-SellNo	6	6	6	11	10	10	10	10	10	10
L10-BidPrice	4936.5	4936.5	4936.5	4936.5	4936.5	4936.5	4936.5	4936.5	4936.5	4936.5
L10-BidSize	370	370	370	370	370	370	370	370	370	370
L10-BuyNo	3	3	3	3	3	3	3	3	3	3
L10-AskPrice	4946.5	4946.5	4946.5	4946	4946	4946	4946	4946	4946	4946
L10-AskSize	1178	1178	1178	1071	1071	1071	1071	1071	1071	1071
L10-SellNo	7	7	7	6	6	6	6	6	6	6

Notes

- One of the main reason is that the proxy is adjusted for the rising trade volumes associated with the period on which the study is made. On the contrary, the period 2005–2016 which is used to construct our sample is not associated with any rising phenomenon of trade volume. Conversely, the LSE has lost market share to the competing alternative trading venues during the same period. See Section 4.2.4 for the detail arguments.
- In order to capture ‘dark pool’ operators and other similar trading systems, a new category of trading venue called Organised Trading Facility (OTF) is introduced for non-equity instruments in MIFID II, which came into effect on 3 January 2018.
- The turnaround time between a message from a trader and its receipt at exchange platform.
- Foucault et al. (2013) define liquidity as the degree to which an order can be executed within a short time frame at a price closer to the security’s consensus value. Conversely, if a price deviates substantially from the consensus value, there is illiquidity.
- The trade signing methodology adopted in this study goes as follows. In a first phase, algorithms filter all trades not sourcing from the automatic session and then accumulate trades executed on the same milliseconds with the same price. The problem arising from accumulating all trades indiscriminately executed in the same millisecond is carefully avoided. Generally, trade records delivered with the same time-stamp include both buy and sell trades. Therefore, it is important to distinguish them as buyer or seller initiated trades before accumulating them. The second phase is bit more complex and time-consuming where algorithms match trade price with the relevant quotes, both bid and ask, considering several “if and then” conditions. The algorithms attempt to match a trade price with the immediately available prior quotes (either bid or sell), if they find a match with bid

then provide a seller-initiated trade flag or a buyer-initiated flag when they find a match with ask. If the algorithms do not find a match with the immediate quotes, then they look for a match to the one before the immediate one, and so on. In contrast, a traditional trade signing approach compares changes in trade price with the changes in mid price to ascertain whether an executed trade is buyer- or seller-initiated, and does not seem to fit a dynamic low-latency environment where quote update speed is very high and the time synchronization between trades and quotes updates is not quite orderly. The algorithms used in this study can assign a trade sign with accuracies reaching over 99%.

⁶ The general expressions for the partial effect are evaluated at the sample means of HFT (96) and market fragmentation (2.17), as reported in Table 4.

⁷ $(0.282 - 0.254) / 0.282 \approx 0.10$, see Table 11.

References

- Ahn, Hee Joon, Jun Cai, and Cheol Won Yang. 2018. Which liquidity proxy measures liquidity best in emerging markets? *Economies* 6: 67. [CrossRef]
- Aitken, Michael, Haoming Chen, and Sean Foley. 2017. The impact of fragmentation, exchange fees and liquidity provision on market quality. *Journal of Empirical Finance* 41: 140–60. [CrossRef]
- Aldridge, Irene. 2013. *High Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading System*, 2nd ed. Hoboken: Wiley.
- AMF. 2017. *Study of the Behaviour of High-Frequency Traders on Euronext Paris*. Technical Report. Available online: <https://www.amf-france.org/en/news-publications/publications/reports-research-and-analysis/study-behaviour-high-frequency-traders-uronext-paris> (accessed on 15 December 2021).
- Baldauf, Markus, and Joshua Mollner. 2021. Trading in Fragmented Markets. *Journal of Financial and Quantitative Analysis* 56: 93–121. [CrossRef]
- Ben Ammar, Imen, and Slaheddine Hellara. 2021. Intraday interactions between high-frequency trading and price efficiency. *Finance Research Letters* 41: 101862. [CrossRef]
- Biais, Bruno, and Thierry Foucault. 2014. HFT and market quality. *Bankers, Markets and Investors* 128: 5–19.
- Boehmer, Ekkehart, Kingsley Fong, and Juan Wu. 2021. Algorithmic Trading and Market Quality: International Evidence. *Journal of Financial and Quantitative Analysis* 56: 2659–88. [CrossRef]
- Boehmer, Ekkehart, Dan Li, and Gideon Saar. 2018. The Competitive Landscape of High-Frequency Trading Firms. *The Review of Financial Studies* 31: 2227–76. [CrossRef]
- Bouveret, Antoine, Cyrille Guillaumie, Carlos Aparicio Roqueiro, Christian Winkler, and Steffen Nauhaus. 2014. High-frequency trading activity in EU equity markets. *European Securities and Markets Authority, Economic Report* 1: 1–31.
- Brogaard, Jonathan, and Corey Garriott. 2018. High-Frequency Trading Competition. *Journal of Financial and Quantitative Analysis* 54: 1469–97. [CrossRef]
- Brogaard, Jonathan, Terrence Hendershott, Stefan Hunt, and Carla Ysusi. 2014. High-frequency trading and the execution costs of institutional investors. *Financial Review* 49: 345–69. [CrossRef]
- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2014. Market Integration and High Frequency Intermediation. Available online: <https://docs.iiroc.ca/DisplayDocument.aspx?DocumentID=89E7649BE79B49A0BAC4B2BE1685EDA0&Language=en> (accessed on 15 December 2021).
- Carrion, Allen. 2013. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets* 16: 680–711. [CrossRef]
- CFA Institute. 2011. The Structure, Regulation, and Transparency of European Equity Markets under MiFID. Technical Report. Available online: <https://www.cfainstitute.org/en/advocacy/policy-positions/structure-regulation-and-transparency-of-european-equity-markets-under-mifid> (accessed on 15 December 2021).
- Conrad, Jennifer, Kevin M. Johnson, and Sunil Wahal. 2003. Institutional trading and alternative trading systems. *Journal of Financial Economics* 70: 99–134. [CrossRef]
- Conrad, Jennifer, Sunil Wahal, and Jin Xiang. 2015. High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics* 116: 271–91. [CrossRef]
- Degryse, Hans, Frank De Jong, and Vincent Van Kervel. 2015. The impact of dark trading and visible fragmentation on market quality. *Review of Finance* 19: 1587–622. [CrossRef]
- Foucault, Thierry, Marco Pagano, and Alisa Roell. 2013. *Market Liquidity: Theory, Evidence, and Policy*. Oxford: Oxford University Press.
- Friederich, Sylvain, and Richard Payne. 2015. Order-to-trade ratios and market liquidity. *Journal of Banking & Finance* 50: 214–23. [CrossRef]
- Frino, Alex, Vito Mollica, Eleonora Monaco, and Riccardo Palumbo. 2017. The effect of algorithmic trading on market liquidity: Evidence around earnings announcements on Borsa Italiana. *Pacific Basin Finance Journal* 45: 82–90. [CrossRef]
- Gomber, Peter, B. Arndt, M. Lutat, and T. Uhle. 2011. High-Frequency Trading. Available online: https://www.deutsche-boerse.com/resource/blob/69642/6bbb6205e6651101288c2a0bfc668c45/data/high-frequency-trading_en.pdf (accessed on 15 December 2021).
- Gresse, Carole. 2017. Effects of lit and dark market fragmentation on liquidity. *Journal of Financial Markets* 35: 1–20. [CrossRef]
- Hagströmer, Björn. 2021. Bias in the effective bid-ask spread. *Journal of Financial Economics* 142: 314–37. [CrossRef]
- Hagströmer, Björn, and Lars Nordén. 2013. The diversity of high-frequency traders. *Journal of Financial Markets* 16: 741–70. [CrossRef]

- Hasbrouck, Joel, and Gideon Saar. 2013. Low-latency trading. *Journal of Financial Markets* 6: 646–79. [CrossRef]
- Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. 2011. Does algorithmic trading improve liquidity? *The Journal of Finance* 66: 1–34. [CrossRef]
- Hendershott, Terrence, and Ryan Riordan. 2013. Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis* 48: 1001–24. [CrossRef]
- Jovanovic, Boyan, and Albert J. Menkveld. 2015. Middlemen in Limit order Markets. Available online: <https://ideas.repec.org/p/red/sed010/955.html> (accessed on 15 December 2021).
- Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun. 2017. The flash crash: High-frequency trading in an electronic market. *Journal of Finance* 72: 967–98. [CrossRef]
- Kollias, Christos, Stephanos Papadamou, and Costas Siriopoulos. 2013. European markets' reactions to exogenous shocks: A high frequency data analysis of the 2005 london bombings. *International Journal of Financial Studies* 1: 154–67. [CrossRef]
- Lee, Charles M. C., and Mark J. Ready. 1991. Inferring trade direction from intraday data. *Journal of Finance* 46: 733–46. [CrossRef]
- Leone, Vitor, and Frank Kwabi. 2019. High frequency trading, price discovery and market efficiency in the FTSE100. *Economics Letters* 181: 174–77. [CrossRef]
- Linton, Oliver, and Soheil Mahmoodzadeh. 2018. Implications of High-Frequency Trading for Security Markets. *Annual Review of Economics* 10: 237–59. [CrossRef]
- Markellos, Raphael N., Terence Mills, and Costas Siriopoulos. 2003. Intradaily behavior of listed and unlisted security basket indices in the emerging greek stock market. *Managerial Finance* 29: 29–54. [CrossRef]
- Menkveld, Albert J. 2016. The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics* 8: 1–24. [CrossRef]
- Mishra, Suchismita, and Le Zhao. 2021. Order Routing Decisions for a Fragmented Market: A Review. *Journal of Risk and Financial Management* 14: 556. [CrossRef]
- O'Hara, Maureen. 2015. High frequency market microstructure. *Journal of Financial Economics* 116: 257–70. [CrossRef]
- O'Hara, Maureen, Chen Yao, and Mao Ye. 2014. What's Not There: Odd Lots and Market Data. *The Journal of Finance* 69: 2199–36. [CrossRef]
- O'Hara, Maureen, and Mao Ye. 2011. Is market fragmentation harming market quality? *Journal of Financial Economics* 100: 459–74. [CrossRef]
- Petersen, Mitchell A., and David Fialkowski. 1994. Posted versus effective spreads. Good prices or bad quotes? *Journal of Financial Economics* 35: 269–92. [CrossRef]
- The Netherlands Authority for the Financial Markets. 2010. High Frequency Trading: The Application of Advanced Trading Technology in the European Marketplace. AFM Report on High Frequency Trading (HFT). Available online: <https://www.afm.nl/~/~profmedia/files/rapporten/2010/hft-report-engels.ashx> (accessed on 20 November 2021).
- The Netherlands Authority for the Financial Markets. 2016. A Case Analysis of Critiques on High Frequency Trading. Technical Report. Available online: <https://www.afm.nl/~/~profmedia/files/rapporten/2016/case-analysis-critiques-high-frequency-trading.ashx> (accessed on 20 November 2021).
- Upson, James, and Robert A. Van Ness. 2017. Multiple markets, algorithmic trading, and market liquidity. *Journal of Financial Markets* 32: 49–68. [CrossRef]
- Zaharudin, Khairul Zharif, Martin R. Young, and Wei-Huei Hsu. 2021. High-frequency trading: Definition, implications, and controversies. *Journal of Economic Surveys* 36: 75–107. [CrossRef]