Finance Research Letters xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

A new approach for detecting high-frequency trading from order and trade data

Cumhur Ekinci*, Oguz Ersan

Istanbul Technical University (ITU), Department of Management Engineering, ITU Isletme Fakultesi, Macka, Istanbul 34367, Turkey

ARTICLE INFO

JEL classification:
G10
G12
G15
G23

Keywords:
High-frequency trading (HFT)
HFT detection
Low latency trading
Borsa Istanbul

ABSTRACT

We suggest a two-step approach in detecting HFT activity from order and trade data. While the first step focuses on multiple actions of an order submitter in low latency, the second searches for the surroundings of these orders to link related orders. On a sample of 2015 data from Borsa Istanbul, we estimate that average HFT involvement is 1.23%. HFT activity is generally higher in large cap stocks (2.88%). Most HFT orders are in the form of very rapidly canceled order submissions. A robustness check reveals a mean accuracy rate of 97% in the linkage of orders.

1. Introduction

High-frequency trading (HFT) is a global phenomenon that has changed the nature of trading and investment in the last decade. In terms of their share in total turnovers and message traffic as well as their impact on the price formation process, high-frequency traders have really intervened in the design of the game all over the world.[1] Various effects of HFT have already been documented in the literature. Its impact on price discovery and efficiency (Carrion, 2013; Brogaard et al., 2014) and liquidity provision (Hasbrouck and Saar, 2013; Menkveld, 2013) as well as its role in intraday price shocks (Brogaard et al., 2016; Kirilenko et al., 2017) are few examples among many others. In order to determine various effects of HFT on financial markets, accurate and straightforward quantification of this activity is essential. Furthermore, considering trading speed as an obvious comparative advantage of HFTs against non-HFTs, precise and broadly applicable detection of HFT is beneficial to the rest of the market participants.

Although the investigation of HFT attracts both academics and professionals, research on this topic is limited primarily due to the lack of appropriate data. Some researchers utilize special datasets provided directly by HFT firms (e.g. Brogaard, 2010; Brogaard et al., 2014; Carrion, 2013; Hirshey, 2016). However there exist at least two drawbacks for this type of datasets. First, they are limited and unique which leaves the question open as to whether the findings are special to the given timespan, market or other study details. An objection is directed towards the market fragmentation issue in Conrad et al. (2015). The authors argue against the exchange-specific results in the case of largely fragmented U.S. market. Second, although the examined dataset is an HFT sample, this does not necessarily represent the whole HFT activity in the market. This fact by itself may generate a bias in drawing conclusions while systematically omitting part of the HFT activity in the analysis (O' Hara et al., 2014).

Another way of capturing HFT activity is through detecting HFT orders from more common data types. However, in this case,

* Corresponding author.
E-mail address: ekincicu@itu.edu.tr (C. Ekinci).

[1] Brogaard et al. (2014) suggest that 68.5% of the dollar volume is generated by HFT which takes part in 74% of the trades in a relatively older dataset from 2008-2009 for the U.S. stock market. Hagströmer and Norden (2013) provide the range of 25% and 50% for the HFT share in total trading activity in Sweden.

http://dx.doi.org/10.1016/j.frl.2017.09.020
Received 11 July 2017; Received in revised form 5 September 2017; Accepted 27 September 2017
1544-6123/

Please cite this article as: Ekinci, C., Finance Research Letters (2017), http://dx.doi.org/10.1016/j.frl.2017.09.020

accurate inference of HFT activity is a challenge by itself. For example, Zhang (2010) attributes all the short-term trading activity (not existing in quarterly holdings reports) of institutions to HFT while Jones (2013) argues against the broad definition of HFT. Kirilenko et al. (2017) classify HFT based on the activity of user accounts. Specifically, accounts with large trading activity and small end-of-day inventory are inferred as HFT accounts. All the trading activity from these accounts are associated with HFT. Conrad et al. (2015) suggest a measure of high-frequency quoting which is any change in best bid or offer (either quote or size) in all venues.[2]

Hasbrouck and Saar (2013) and Ersan and Ekinci (2016) employ a more detailed but yet commonly available data type, i.e. electronic order and trade messages in detecting HFT activity. The former focuses on a specific HFT strategy where cancelled orders are shortly followed by orders of same size and direction. In this way, 54% to 60% of all the cancellations are attributed to HFT. The suggested HFT measure is shown to have correlation of around 80% with HFT firms' activity. The latter extends Hasbrouck and Saar by considering also for orders which are submitted, modified and cancelled simultaneously.

In this study, we develop a novel methodology to detect HFT activity from the electronic messages data. We start from low latency reaction as a distinguishing HFT feature in a two-step detection method. By detecting the most obvious HFT orders at the first step and widening this definition by forming links around these orders, we seek high accuracy. On a sample of 422 stocks listed in Borsa Istanbul (BIST), we find that as of 2015, average HFT level is 1.23%. Moreover, HFT activity is concentrated in relatively small number of stocks: mainly largest cap stocks (2.88%). On the other hand, HFT interest in small cap stocks is substantially low (0.20%). We provide robustness checks for the precision of our HFT measure. For around 97% of linked orders, the institution that submits the linked orders is unique.

The remainder of the paper is organized as follows. Section 2 briefly describes our data and the financial market characteristics. Section 3 explains our methodology for detecting HFT orders within the whole order and trade data while the next section provides descriptive results on HFT level proxied by the use of this methodology. Section 5 reports on the accuracy and advocates the use of our proxy. Final section concludes.

2. Data

We study all the listed stocks in Borsa Istanbul (BIST) for the five-month period from June to November 2015. Among 422 examined stocks listed in BIST All index through the study period, we perform most of our analyses for the stocks listed in BIST 30 and BIST 100 indices separately. BIST 30 and BIST 100 indices are constructed based on the turnover and market capitalization of stocks. BIST 30 index can be considered as the index for the blue chip stocks while BIST 100 refers to a broader market. The stocks which are not listed in BIST 100 index are relatively much smaller in terms of their share in total trading activity. Specifically, BIST 30 and BIST 100 stocks are responsible for 72% and 92% of overall turnover through the five months period of our concentration, respectively.

We work with three monthly datasets i.e. a file that stores all order submissions, another file that comprises cancellation and modification messages, and a final one for trades. Each file contains descriptive information such as size, price, order ID, message number reflecting chronology and message type (order submission, modification, cancellation or execution) for each electronic message. We combine these three files on daily basis and obtain chronological sequence of electronic messages for each order in the stocks.[3] All orders have a sequence of messages that starts with a submission message and ends with an execution or a cancellation.

In our dataset, time stamps are given in seconds which is a relatively long interval for an HFT analysis. However, overall trading activity as well as HFT and order submission speed are much lower when compared to developed financial markets.[4]

Table 1 reports statistics on trade intervals in Borsa Istanbul. The statistics except the last row are on the stock level, utilizing intervals between trades in one stock at a time. Only 1.53% of all trades follow trades in same stock within same second. For small stocks, this level is even lower at 1.34%. Average median trade interval for the stocks listed in BIST All is as high as 29.68 s while this reduces to 11.93 s for BIST 30 stocks. Thus, we expect not to lose much value in relying on timestamps being reported in seconds.

The purpose of selecting the specified time span primarily is to use the additional information (broker IDs) provided during this period. On Nov 30, 2015, Borsa Istanbul upgraded its trading technology (the new one called BISTECH) and changed its data format. Before this very date, IDs of brokers through which orders are submitted were released. Hence, we are able to use these additional data in order to test the accuracy of our method.

Table 2 provides descriptive statistics. Within 126 trading days and for 422 stocks, we have roughly 85 million submitted electronic messages. Among these, 67.9 million are order submissions while the remaining are modifications and cancellations. After combining partial executions of a trade in one, we have 25.8 million "unique" trades. On the stock level, largest number of daily orders (trades) is as high as 94,348 (46,844). Moreover, there exist daily 13,437 modifications at a specific stock. Overall fill ratio is 57% which signals a notable change when compared to 67% for a 2013–2014 dataset documented in Ersan and Ekinci (2016).

Borsa Istanbul is an emerging market with $1.6 billion daily turnover by June 2015. All publicly held companies are listed in BIST. Except one stock (Turkcell Co. which is also NYSE listed), there is no market fragmentation. There are two trading sessions: morning and afternoon. Trading is continuous through 09:35–12:30 and 14:15–17:30. There are three call auctions. Prices are fixed

[2] See Table A1 in Aitken et al. (2015) for a relevant list of studies identifying HFT activity.

[3] Since the order validity was no longer than one day in BIST stocks, we did not need to go to previous day's data to calculate current day's HFT.

[4] Our own tests in two different days in 2017 (Feb 10 and Aug 23) through a traditional channel for order submission (off-the-colocation) revealed that there is around 0.3 s time interval between the execution of two basket orders submitted simultaneously. In case of a conditional order (e.g. take-profit), this interval reaches 0.6 s. This shows that although algorithmic traders can seek low latency, for a typical non HFT trader, it is very unlikely to get into trading activity within the same second. Besides, the fact that we work with order ID numbers limits any confusion.