Cross-Venue Liquidity Provision: High Frequency Trading and Ghost Liquidity

Hans Degryse, Rudy de Winne, Carole Gresse, Richard Payne

Cross-Venue Liquidity Provision: High Frequency Trading and Ghost Liquidity*

Hans Degryse[a], Rudy De Winne[b], Carole Gresse[c], and Richard Payne[d]

[a] *KU Leuven, CEPR*
hans.degryse@kuleuven.be

[b] *UCLouvain*, *Louvain Finance (LIDAM)*
rudy.dewinne@uclouvain.be

[c] *Université Paris Dauphine-PSL, DRM, CNRS*
carole.gresse@dauphine.psl.eu

[d] *Cass Business School*, *City University of London*
richard.payne.1@city.ac.uk

July 2, 2020

JEL classification: G14, G15, G18

Keywords: High Frequency Trading (HFT), Algorithmic Trading (AT), Fragmentation, Ghost liquidity

* We would like to thank Panos Anagnostidis, Carlos Aparicio Roqueiro, Jos van Bommel, Antoine Bouveret, Carole Comerton-Forde, Sarah Draus, Monika Gehde-Trapp, Cyrille Guillaumie, Björn Hägstromer, Frank Hatheway, Peter Hoffmann, Charles Jones, Olga Klein, Christophe Majois, Katya Malinova, Giang Nguyen, Carol Osler, Christine Parlour, Tavy Ronen, Gideon Saar, Anne-Laure Samson, Duane Seppi, Andriy Shkilko, Elvira Sojli, Ingrid Werner, Christian Winkler, Gunther Wuyts as well as participants of the Group of Economic Advisors at ESMA, seminar participants at KU Leuven, participants at the 35th Spring AFFI conference, the 4th WiM meeting, the 2nd ECM Workshop, the 2nd SAFE Market Microstructure Conference, and the 46th EFA meeting (Lisbon) for their comments. We thank Yujuan Zhang for excellent research assistance, ESMA for providing access to the data, and the French National Research Agency (ANR) for funding through project GHOST. The views presented in this paper are those of the authors and do not necessarily reflect the views of ESMA.

Cross-Venue Liquidity Provision: High Frequency Trading and Ghost Liquidity

Abstract

We measure the extent to which consolidated liquidity in modern fragmented equity markets overstates true liquidity due to a phenomenon that we call Ghost Liquidity (GL). GL exists when traders place duplicate limit orders on competing venues, intending for only one of the orders to execute, and when one does execute, duplicates are cancelled. Employing data from 2013 for 91 stocks trading on their primary exchanges and three alternative platforms where order submitters are identified consistently across venues, we find that simply measured consolidated liquidity exceeds true consolidated liquidity due to the existence of GL. On average, for every 100 shares passively traded by a multi-market liquidity supplier on a given venue, around 19 shares are immediately cancelled by the same liquidity supplier on a different venue. Yet the average weight of GL in total consolidated depth, at around 4%, does not outweigh the liquidity benefits of fragmentation. GL is most pronounced for traders with a speed advantage such as high-frequency traders, in stocks exhibiting greater market fragmentation, in stocks where the tick is more likely to be binding, and on non-primary exchanges. Furthermore, GL decreases when the fraction of traders using smart order routing is large. Finally, we show that an increase in GL leads to the execution costs of slow and algo traders increasing, while those of HFTs are unaffected.

1. Introduction

The ability to accurately measure liquidity in financial markets is crucial both for traders who want to formulate an optimal execution strategy and for regulators who wish to assess the quality of operation of financial markets. However, recent developments in market structure have made this measurement task difficult. First, the fragmentation of modern equity markets and the use of multiple trading venues by market participants means that to measure liquidity one must aggregate across many venues and data feeds to obtain a 'consolidated' view of the market, while to execute efficiently requires the use of a 'smart order router' (see, for example, Foucault and Menkveld, 2008). Second, though, the same market developments have led to changes in traders' order submission strategies which imply that 'consolidated' liquidity (measured as the simple aggregate of shares available across all trading venues) is likely to be an overstatement of the actual liquidity that an average trader can access. We define 'Ghost Liquidity' (GL) to be the magnitude of this overstatement.

To understand GL, consider a simple scenario in which all participants involved in trading a stock have access to two venues. A patient investor who wishes to buy a unit of the stock might place a limit buy order on one of the two venues. She then executes if a matching market sell arrives at this venue. However, she misses out on trading opportunities if market sells are arriving at the other venue. Thus, to maximize her chances of execution, she is incentivized to place similar limit buy orders on both venues and intends, when one of the orders has executed, to cancel the other. It is this order duplication that is at the heart of what we call GL. Let us imagine that an impatient but unsophisticated trader places a market sell order to hit the limit buy order posted on one of the two venues but that, at the same time, the same limit buyer is executed on the other venue. If the seller's trading technology is slower than that of the patient buyer, by the time her sell order reaches the market, the limit buy order she targets will have been cancelled. As a result, the liquidity actually accessible to her is less than initially observed, and this difference is our definition of GL.

Of course, order duplication is not without risk. If both of our passive trader's limit buy orders are hit simultaneously, she will have executed too great a quantity. This double execution may occur either because the duplicated orders are hit at each venue by two different traders or because a single trader using a smart order router intentionally executes the passive trader's orders on both

venues simultaneously. This simple example implies that the incentive to duplicate limit orders across venues is greater for traders who have a trading speed advantage over the average trader but the incentive is weakened by the presence of Smart Order Routers (SORs).

In brief, in a world of fragmented trading, the replication of orders across venues by fast traders leads measured liquidity to overstate true liquidity for the average trader. To be clear, we are not defining GL to arise from orders which were never intended to execute under any circumstance (which may also be a problem in modern markets), but to arise due to orders which are cancelled conditional on an order submitted by the same trader being filled on another venue. This means that GL is not necessarily inaccessible to all traders. It is an unstable form of liquidity that turns out to be unavailable to the average trader most of the time.

The core of this paper is an attempt to quantify the size of GL in equity markets and to characterize its determinants. We take advantage of a unique data set that covers 91 European stocks trading on their respective primary exchanges and the three largest alternative European trading venues for the month of May 2013. The data contain the usual order level and individual trade information that is common to many modern microstructure databases, but importantly the data also provide anonymized information on the market members who submitted each order. Thus we can track market members across time, across stocks, and across trading venues. This identity information can also be used to characterize those participants in terms of trading speed and technology.

With these data we measure GL by computing a trader's voluntary cancellations of liquidity on one venue following execution of one of that trader's similar orders on another venue. Then we aggregate across traders, venues, and time to assess the overall size of GL as a fraction of the size of the triggering execution, and also as a fraction of total liquidity, and we regress GL measures on a set of trader characteristics, venue characteristics, and exogenous variables to characterize its determination.

We find that GL accounts for a sizeable fraction of order cancellation activity. To a rough approximation, execution of one of the average participant's limit orders on a particular venue, leads her to cancel quantity equivalent to roughly 20% of the size of that trade on the other venues where she has posted similar orders. There are variations across venues and countries, with GL measured as a percentage of trade size ranging between less than 1% in Spain and over 40% in the UK (i.e., in UK stocks, following a trade on one venue, shares equating to 42% of the original

trade size are cancelled on a competing venue on average). GL is larger for stocks with greater market capitalization, it is smaller in more volatile markets and, as one might expect, it is considerably greater for stocks with a large degree of fragmentation.

Our investigation of the determinants of GL also shows that trader characteristics are important. High-frequency traders (HFTs) have the largest measures of GL, followed by algorithmic trading (AT) firms. Traders who are neither HFTs nor ATs, a group that we call 'slow' traders, have the lowest GL levels. GL is also larger when a trader is acting as a principal rather than as an agent.

Using a Tobit analysis for data measured at a 15-minute interval, we find that, in addition to the results above, traders tend to use ghost orders most heavily when other traders are also doing so and that GL increases when stock-specific trading volumes are high. We also observe that when the execution that triggered the ghost liquidity removal was large, the fraction of displayed liquidity that a trader removes increases. There is also evidence that GL effects are strongest when the triggering trade is on an alternative trading venue (i.e., not the primary exchange in a country) and when the venue where liquidity is being removed is also an alternative trading venue.[1] This implies that alternative venues, perhaps because of the trader population that they attract and the low latencies that they offer, are more affected by GL. Finally, we find that when the (absolute) inventory level of the liquidity supplier is large, (s)he is marginally less likely to duplicate orders. This leads us to reject our hypothesis that GL is used as a tool for rebalancing excessive inventories. Instead, order submissions to multiple venues might be used as a strategy to build inventory. We also find that when the prevalence of smart order routing is particularly large, smart order routing tends to reduce ghost liquidity. Again we expect that this effect is due to the likelihood of multiple executions when smart order routers are a significant factor in the market.

We proceed from the analysis of the determinants of GL to a study of its implications. We examine whether the level of GL impacts upon the daily execution costs, measured by effective spreads, that various trader groups pay. We find that slow traders' execution costs increase with GL posted on the primary trading venue for a stock while algorithmic traders' execution costs increase with GL on all venues. Thus the use of order duplication and subsequent cancellation renders less sophisticated traders' execution strategies less effective. The fact that GL on the primary venue matters for slow traders is consistent with the fact that slow traders do the vast

[1] Examples of primary exchanges are the London Stock Exchange and Euronext Paris, while our alternative trading venues are BATS, Chi-X and Turquoise.

majority of their trading on the primary venues. It is worth noting that HFT execution costs do not suffer from GL as, presumably, their sophistication and trading speed insulate them from its effects.

Thus, overall our results show GL to be an economically significant phenomenon. Measured liquidity and 'true' liquidity can differ substantially especially for stocks with high HFT activity and large fragmentation. This raises questions about the use of simple consolidated liquidity measures to assess market quality and to measure the effects of changes in regulation. Furthermore, the result that higher GL is associated with greater execution costs for less sophisticated traders is also concerning from a regulatory perspective.

The rest of the paper is structured as follows. Section 2 contains a brief overview of relevant literature. Section 3 is an introduction to our data. Section 4 gives a description of how we classify market participants using our data and Section 5 presents our initial measurements of GL. Section 6 contains our analysis of the determinants of GL. We examine the impact of GL on trading costs in Section 7, and Section 8 provides some conclusions from our work.

2. Literature review and research objectives

We are interested in measuring and characterizing the determinants of ghost liquidity (GL). By GL, we mean liquidity that is supplied to markets but which is not intended to execute (or perhaps not intended to execute in full). This could occur in a single consolidated market, with a trader submitting multiple buy or sell orders to different levels of an order book (in order to gain time priority), only one of which is intended to execute. It could also occur in fragmented markets, where duplicate orders in the same stock are sent to many venues. What are the incentives to post multiple orders? In fragmented markets, for illustration, duplicating orders across venues allows one to avoid time priority and increases one's chances of execution if there are aggressive traders who operate only on single venues. However, this increased execution probability is not without risk, as there is the chance that two or more of the duplicate orders are filled, leading to over-trading. Regardless, in all cases GL is likely to be characterized by (i) over-supply of liquidity relative to true trading intentions and, as a consequence of this, (ii) cancellation of the excess supply of limit orders once one of them executes.

Recent literature has demonstrated that there may be over-supply of depth on a single venue, resulting from the imposition of time priority and variations of trading speed across participants.

Yueshen (2014), for example, argues that following changes in asset prices, there may be a race by fast traders to be the first-in-line at the new equilibrium price leading to a temporary spike in depth before traders realize their actual position in the queue and, through subsequent cancellations, depth normalizes. Blocher et al. (2016) identify clusters of extremely high and extremely low limit order cancellation activity using data on all the S&P 500 stocks for the calendar year of 2012. They find that cancel clusters largely appear to be generated by HFTs sparring with one another to get to the front of the limit order queue, rather than HFTs trapping unsuspecting investors into bad executions. Dahlström et al. (2018) investigate the economic rationale behind limit order cancellations from the perspective of liquidity suppliers. They show that changes in common values affect the value of a limit order depending upon the queue position, but HFTs behave in a similar way as other traders. These papers suggest that competition between fast traders on the same venue can lead to 'excess' depth in the short-run that is eliminated by cancellation activity. Dahlström et al. (2018) further show that trades at competing venues lead to significant cancellations at the primary venue; the economic significance of this force relative to other determinants of cancellations however is low.

GL may also arise due to fragmentation in trading across venues. Fragmentation has been an important feature of equity markets since the early 2000s in the U.S. and since the introduction of MiFID trading rules in 2007 in Europe. Traders who are connected to many competing trading venues can benefit by accessing the separate liquidity pools on those venues. Empirical research indicates a strong link between fragmentation and measured liquidity. Foucault and Menkveld (2008) show that, due to the absence of time priority across markets, consolidated depth is larger after the entry of a new order book. O'Hara and Ye (2011) find that, for U.S. stocks, spreads are tighter and price efficiency is higher with fragmentation. Degryse, de Jong and van Kervel (2015) find that lit fragmentation (i.e., fragmentation across pre-trade transparent venues) in Dutch stocks has increased liquidity through reductions in bid-ask spreads and increases in depth across markets. Gresse (2017) employs data for stocks listed on the London Stock Exchange (LSE) and Euronext and finds that lit fragmentation improves bid-ask spreads and depth across markets.

However, while the results above clearly indicate that fragmentation leads to larger *measured, consolidated* liquidity, it is possible that *measured* and *real* liquidity differ. If investors cannot tap all depth at all venues simultaneously, they cannot benefit from the greater consolidated liquidity. This may occur for at least two reasons.

1) Some investors may lack the technology to connect to several venues and therefore be restricted to accessing the primary exchange only. Degryse et al. (2015) and Gresse (2017), for example, show that the benefits of fragmentation are not accessible to investors who are restricted to accessing the primary exchange only.
2) Fast order cancellations may alter the true level of depth. Hasbrouck and Saar (2009), for instance, identify trading strategies that involve 'fleeting orders' which are orders that are submitted then cancelled very rapidly. If liquidity suppliers have a latency advantage, then their speed of cancellation may mean that the depth on an order book is difficult to access for a slow liquidity demander. In such a setting, suppliers may post duplicate limit orders on more than one venue, only intending for one of the orders to execute and cancelling the duplicates once an execution occurs. The latency advantage enjoyed by liquidity suppliers means that they face limited asymmetric information risk and that the risk of being over-filled is small. It is this order duplication across venues that we define as GL and which implies that measured, consolidated liquidity is larger than real liquidity.

Our definition of GL above suggests that looking at order duplication and order cancellations on one venue in response to trades on another might be useful in identifying GL. This approach is used in ESMA (2016), who use the same data as we do to show that around 20% of all limit orders are duplicated, with the duplication strategy used more frequently by HFTs and for large cap stocks.[2] They also show that following around a quarter of all trades, the liquidity supplier cancels duplicate orders on other trading venues. Chen et al. (2017) do not study duplication but focus on cancellations and their implications for the difference between measured and real consolidated liquidity in fragmented markets when there are latency differentials between traders. They study the introduction of an asymmetric, randomized speed bump to the Canadian exchange TSX Alpha on September 21, 2015, which low-latency traders could avoid by paying a fee. After the introduction of the speed bump, low-latency liquidity providers on Alpha are shown to use their speed advantage to cancel delay-exempt limit orders and thus "fade away" from incoming market orders which consume liquidity from multiple venues. Thus displayed depth overstates real depth. The results also imply that existing empirical findings on the benefits of fragmentation may be flawed.

[2] We contributed to the development of the measures used in the ESMA report as independent experts.

However van Kervel (2015) shows that in worlds with no GL (by our definition) one might also observe such cross-venue cancellations in response to trades on other venues. He builds a model with multiple venues and where HFT market-makers post quotes on all venues simultaneously. In the absence of any new information those market-makers would be willing to trade at those quotes on all venues and would not choose, for example, to cancel or modify quotes on venue B in response to a trade on venue A. In this sense, those quotes are real and not ghost. However, if there is asymmetric information then a trade on venue A will lead to quote updating (through cancellations and modifications) on all other venues. Again, this is not because the original quotes were 'phantom liquidity', it is rational updating of quotes in response to new information. Thus one observes cross-venue cancels in a world without GL. Employing data from the LSE and four competing exchanges, van Kervel (2015) finds that once a market order consumes liquidity on one venue, the depth available at other venues is reduced. Two takeaways from van Kervel's work are that (1) it will be important for us to account for asymmetric information effects if we want to understand cancellation activity; and (2) estimates of GL simply based on cancellations, without tracking traders individually, would be biased as those cancellations might reflect the rational updating of dealers' quotes in response to information revealed by trades.

One key issue in identifying the importance of GL is that one needs to be able to track the same traders across venues. The observed drop in depth on other venues after a trade on one venue could simply capture the equilibrium responses of all traders to the trade event. Our research overcomes this identification challenge by following the same traders across venues.[3] We are therefore able to make four important contributions to the literature. First, we estimate the importance of GL for a given trader. Second, we compare the importance of GL across different groups of traders, and across different venues. Third, based on our measurement of GL by trader, we identify economic determinants of GL. Last, we assess the impact of GL on the execution costs of different groups of traders.

3. Sample, data, and market organization

We employ a proprietary dataset collected by ESMA and several National Competent Authorities for the month of May 2013. It consists of 91 stocks that are primary listed on the

[3] It is worth noting that the other literature, mentioned above, cannot track individual traders across venues in their data.

historically main exchanges of nine countries comprising Belgium, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, and the United Kingdom. The dataset covers the primary exchanges[4] and trading/quoting activity on the three largest alternative exchanges in action at that time, namely BATS, Chi-X, and Turquoise, which together represent the vast majority of trading activity for each stock. ESMA (2014) were the first to employ the data set, in their analysis of the extent of HFT in European stock markets. Further details on the construction and content of the data set can be found there.

All exchanges in our study are regulated under the Markets in Financial Instruments Directive (MiFID). The national exchanges where our sample stocks are primary listed will be referred to as "primary" exchanges and denoted *PE*, and other trading venues where the stocks are admitted to trading will be referred to as "alternative" exchanges and denoted *ALT*.

In terms of market organization, all trading platforms considered in our study operate as open, transparent, and anonymous electronic order books on which buy and sell orders are continuously matched from the open to the close according to price/time priority rules. Primary exchanges commence and finish their trading sessions with call auctions while no call auctions are organized on alternative venues either at the open or at the close. Further, alternative venues use a make/take fee structure that remunerates liquidity-providing orders and charges aggressive orders.

The set of stocks in the sample was built using a stratified sampling approach taking into consideration market capitalization, value traded, and fragmentation. For each country, stocks were split by quartiles according to their market value, value traded, and their level of fragmentation across venues, using December 2012 data. A random draw was performed to select stocks in each quartile. In order to account for the relative size of the markets, greater weight was put on larger countries. At the same time, at least five different stocks were selected from each country. This procedure yielded an original sample of 100 stocks from which nine stocks had to be excluded due to thin trading issues.[5] As a result, the number of stocks in two of our sample countries fell to just four. The final sample includes stocks with very different features. The average daily value traded ranged from less than EUR 0.1mn to EUR 611mn. In terms of market

[4] The primary exchanges are Euronext Amsterdam, Euronext Brussels, Euronext Lisbon, Euronext Paris, Deutsche Börse, Borsa Italiana, the London Stock Exchange, the Irish Stock Exchange, and the Spanish Stock Exchange.
[5] Either those stocks were not traded over several days or they were not traded outside the primary exchange.

capitalization, values ranged from EUR 18mn to EUR 122bn. The breakdown of stocks per country and descriptive statistics for those stocks are provided in Table 1.

Table 1 about here

The entire dataset includes around 10.5 million trades and 456 million messages. Message types include transactions plus order entries, modifications, and cancellations. The unique feature of the dataset is that it contains information on the identity of the market participant behind each message allowing us (i) to follow a market participant across trading venues, and (ii) categorize each participant as an HFT or non-HFT. There is also a capacity flag for each event which indicates whether the member in question is acting in a proprietary or agency capacity.

4. Market member identification and classification

The ESMA dataset contains the list of all market members active on each trading venue during May 2013. There are 388 members in total for our 91 sample stocks. For each message in the dataset, those market participants are identified by anonymized member IDs at several levels of granularity. First, each account for a particular member on a given venue is identified by a specific ID, which we call the Unique ID. Second, all accounts of a given member on a given venue are identified with a common venue-specific ID, designated as the Account ID. Last, if a market participant is a member of several venues, all the accounts of that member are identified on all venues with a common cross-venue ID, designated as the Group ID. This Group ID allows us to follow a market participant across venues. In addition, the dataset provides information about member capacities. For each message, a flag indicates whether the member submitted the message as principal or agent.

From there, we establish and use three member classifications: (1) a slow/fast trader classification based on the HFT identification established by ESMA, (2) a distinction between local members, that is members acting on a single venue, and global members, that is members trading across venues, and (3) a liquidity supplier/taker distinction.

4.1. Slow/fast trader identification

According to MiFID II (cf. Article 4(1)(40)), an HFT technique is "an algorithmic trading technique characterized by: (a) infrastructure intended to minimize network and other types of latencies, including at least one of the following facilities for algorithmic order entry: co-location,

proximity hosting or high-speed direct electronic access; (b) system-determination of order initiation, generation, routing or execution without human intervention for individual trades or orders; and (c) high message intraday rates which constitute orders, quotes or cancellations". As HFT is a rather recent phenomenon, the definitions are still evolving and the academic literature contains many approaches to classify market participants as HFTs or non-HFTs but none of them is perfect.

Two main approaches are often used and sometimes combined. First, firms may be classified as HFT or non-HFT firms based on public information available about their primary business and the types of algorithms or services they use. This approach will be referred to as the direct approach. Second, an analysis of firms' trading strategies (e.g., order placement and cancellation) can also allow a researcher to identify HFTs and we refer to this as the indirect approach. HFT strategies are often characterized by a very short order lifetime (Hasbrouck and Saar, 2013), a high order-to-trade ratio (Hendershott et al., 2011), and an inventory management policy that leads to traders carrying no significant positions over-night (Jovanovic and Menkveld, 2016; Kirilenko et al., 2016). In the search for a more precise HFT classification, these criteria are sometimes combined. For example, Brogaard et al. (2014) and Carrion (2013) use a NASDAQ dataset that includes information on whether the liquidity demanding order and liquidity supplying side of each trade is from an HFT. In their data, Nasdaq defined a firm as an HFT based on both the quantitative properties of that firm's order submissions and trading behavior and on more general information on the firm's business model. But as mentioned by these authors, this combination of criteria and approaches does not allow for a perfect identification.

Our approach to categorizing firms by speed consists of two steps. First, we identify the set of fast traders using the indirect approach of ESMA (2014) based on the lifetime of orders.[6] Second, we identify a subset of fast traders as HFTs using a direct approach.

Bouveret et al. (2014) use the same data as we use with the objective of measuring the extent of HFT in European stock markets. We employ their indirect approach which classifies members as fast traders if the 10% quickest order modifications and cancellations in a given stock occur no more than 100ms after the initial submission.[7] Such a criterion indicates that the member under

[6] We contributed to the preparation of this report as independent experts.

[7] 100ms is clearly below human reaction time. For purposes of comparison, the average duration for a single blink of a human eye is 0.1 to 0.4 seconds, or 100 to 400 milliseconds, according to the Harvard Database of Useful Biological Numbers.

consideration possesses fast trading technology even if she does not use it at all times. It is worth noting that ESMA (2014) find that just over 40% of value traded is done by fast traders using this approach. They also do some robustness checks, varying the 100ms threshold, and show that, while fast trading intensity and the threshold are obviously positively related, the slope of the relationship is fairly flat between 50ms and 250ms. As discussed in ESMA (2014), we choose a fast trader identification based on the lifetime of orders because our main concern is trading speed, regardless of trading strategy. Criteria based on inventory management may identify fast traders implementing market-making strategies but not necessarily other fast traders. An identification based on order-to-trade ratios could also be biased as slow traders with very few trades could be wrongly identified as fast.

The fast trader flag is established by Group ID, by capacity (agent or principal), and by stock. Therefore, a member may be a fast trader for some stocks and not for others, and for a given stock, a member may be considered as a fast trader when trading as principal but not when trading as agent. However, if a given market participant is considered as a fast trader for his proprietary activity in stock *i* on venue *v*, he will be flagged the same way for his proprietary activity on the other trading venues.

We then subdivide the population of fast traders into two categories. We use ESMA's direct approach to identify a list of 21 HFT firms (see ESMA, 2014). This list is built using firms' websites and the financial press to identity each firm's primary business, the use of services to minimize latency, and membership of the European Principal Trader Association. Any fast trading firm that is on this list and is trading as principal is defined as an HFT. We define algorithmic traders (ATs) as the set of participants using computer-based trading technology who are not previously identified as HFTs. These firms are essentially investment banks. In common usage, algorithmic trading is any type of computer-based trading including HFT. In our paper, for clarity, ATs and HFTs are two non-overlapping groups of fast traders.

4.2. Global/local member identification

Not all market participants are active on multiple venues during our sample period. Of the 388, 307 trade on only one venue (with 297 trading only on the primary exchange, 8 trading only on Chi-X and 2 only on Turquoise). There are 39 members who trade on all four platforms, 17 trade on three platforms only, and 25 trade on two platforms only. Thus, in total, 81 members trade on multiple platforms. The 39 market participants trading on all venues account for about 71% of all

trading volume. 20 of the 39 are in the top 10% of market participants as measured by total trading activity. The 307 single-market players represent about 18% of total trading volume in our dataset. Most of them typically trade only a few stocks, but 11 of the 307 are in the top 10% of market participants by activity.

The distinction between members trading at several locations, hereafter called global members, and members trading in a single market, hereafter referred to as local members, is instrumental to our study as GL is defined as a side effect of multi-market trading strategies. We therefore classify global members as market participants who trade in at least two markets and execute more than 10% of their trading volume away from their main trading venue. Any member trading more than 90% of their volume in one market is classified as a local member. This classification is established by Group ID, capacity, and stock.

4.3. Liquidity supplier/taker identification

GL is the outcome of trading strategies in which liquidity is offered at several locations in order to minimize non-execution risk or, equivalently, to capture fragmented market order flow. As such, GL can only be generated by traders implementing passive (i.e., limit order based) strategies. For that reason, it seems relevant to us to distinguish members who are mainly passive in their trading strategies from those who are mainly active. The former will be referred to as liquidity suppliers (LS) and the latter will be referred to as liquidity takers (LT). A member is considered as an LS (LT) if she is the passive (active) counterpart in more than 50% of her total consolidated trading volume when trading as principal. Finally, it is important to note that any member trading as agent is always considered a LT, as agents are executing position changes on behalf of clients rather than taking the other side of public orders and thus seeing their own account affected. This classification is again established by member, by capacity, and on a stock-by-stock basis.

4.4. Member combined classification

A particular member in our data may engage in both principal and agency trading. Where a member in a given stock engages in both, these activities are separated in the data set via the previously mentioned capacity flag, resulting in distinct member/capacity pairings for that member and that stock. While ESMA (2014) argue that the capacity flag cannot be used without difficulties to identify HFTs when using a direct approach and looking across stocks, the capacity flag can still be used for analysis at the stock level. The AT, HFT, global, and liquidity supplier flags are then assigned to each member/capacity pairing, on a stock by stock basis. As a result, the

classification applied to our 388 members produces 8,568 triplets of member×capacity×stock combinations. Further, for the sake of simplicity, in the remainder of the paper when we use the term 'member' 'or trader' we mean a member/capacity pairing.

The scheme described above generates 16 categories of traders (i.e., principal versus agent, slow trader versus AT or HFT, liquidity supplier versus liquidity taker, and local versus global). These are presented in Table 2, along with the number of member×capacity×stock combinations that falls into each category plus their market shares in trading. Note that there are 16, not 24, categories as those trading as agents are never classified either as liquidity suppliers or as HFTs.

Table 2 about here

The largest subgroups correspond to slow local liquidity takers trading as agent (38.0% of member×capacity×stock triplets) and slow local liquidity takers trading as principal (14.5%). Fast traders (i.e., ATs and HFTs), global traders, and liquidity suppliers represent respectively 20.3%, 34.5%, and 18.8% of the population, with fast global liquidity suppliers representing 5.2% equally distributed between ATs and HFTs.

In terms of trading volumes, Table 2 shows that 64.35% of the total volume is traded on primary exchanges while Chi-X is the main alternative venue with 20.91%. ATs and HFT firms account for respectively 22.98% and 22.21% of the total traded value. Their relative weight is greater on BATS, Chi-X, and Turquoise, where the respective volume shares of ATs and HFTs are 26.40% and 32.47%. Trading volume from members trading as principal accounts for 74% of the total volume and is distributed equally between slow and fast traders. Global traders account for 72.81% of total traded volumes and for 96.02% of the volumes traded on alternative venues. Since a local member is defined as a member trading more than 90% of its volume on one venue (often the primary exchange), the very small percentages of volumes observed for local traders on alternative venues are to be expected. Lastly, liquidity suppliers account for 25.47% of the total traded value. They are relatively more active on alternative venues, where they trade 37.45% of the volumes.

5. Assessing the level of ghost liquidity (GL)

As mentioned in Section 4, the Group ID available in our database allows us to follow any market participant across venues. This makes it possible to estimate the amount of GL at different levels of aggregation (trader, venue, …). Subsection 5.1 describes the methodology we use to

measure GL and to aggregate it at different levels. Subsection 5.2 describes how we check whether the GL we measure is actually fictional depth or whether it is immediately followed by re-supply of liquidity by the same trader but at a different price point, thereby reflecting quote updating. Subsection 5.3 reports descriptive statistics.

5.1. Measuring GL

Our GL metric is based on the following simple intuition. Assume that a trader is posting limit sell orders, for example, on several venues simultaneously. Assume also that at a certain time the limit order on the first venue is executed. If, after the execution of the order on the first venue, the trader's limit orders on other venues are left in their respective order books then those orders constitute real liquidity. If, on the other hand, when the order on the first venue executes, the limit orders on other venues are swiftly cancelled then those cancelled orders represented GL.

As the simple example above makes clear, GL has many dimensions. It is trader specific and it might be venue specific. Also, there are several parameters to be specified. How quickly does a trader's order have to be cancelled in response to an execution of another of that trader's orders on a different venue to qualify as GL? How similar does the cancelled order have to be to the executed order to count as GL? Any definition of GL will have to be flexible enough to take account of all of the above.

We begin with a specification of GL as follows. Assume that at time τ a limit sell order posted by member *m* for stock *i* was executed on venue *tv*, the trade venue, and that member *m* had also posted a limit sell order for stock *i* on venue *qv*, the quote venue. Then the sell-side GL posted by *m* on venue *qv* is equal to:

$$GL^{ask}_{tv\to qv}(\tau;\Delta\tau;i;m) = PREQTY^{ask}_{qv}(\tau;i;m) - POSTQTY^{ask}_{qv}(\tau;\Delta\tau;i;m) - \sum_{\tau;\Delta\tau} Volume^{buy}_{qv}(i;m) \quad (1)$$

where $PREQTY^{ask}_{qv}(\tau;i;m)$ is the total limit sell order quantity posted by trader *m* on venue *qv* at the last order book snapshot prior to the trade executed on venue *tv* and $POSTQTY^{ask}_{qv}(\tau;\Delta\tau;i;m)$ is the total limit sell order quantity posted by member *m* on venue *qv* at the order book snapshot that is exactly $\Delta\tau$ seconds after the original snapshot. Thus, the first pair of terms on the right-hand side of the definition measures the reduction in quantity posted by trader *m* on venue *qv* over a small time window (i.e., $\Delta\tau$) around the time of the trade on venue *tv*. The final term on the right-hand side consists of all executions against trader *m*'s limit sell orders on venue *qv* in that same

window. $Volume_{qv}^{buy}(i;m)$ is defined as the size of a market buy order, executing against one of market member *m*'s orders on venue *qv* for stock *i* at any time within the time window. So, all that this definition does is to take the change in total quantity offered by trader *m* and deduct that part of the change that is due to execution activity. The remainder represents voluntary reduction in limit order provision on venue *qv* after the trade on venue *tv* and we count this as GL.

As order book snapshots have been built every 10 milliseconds in the database, the time interval over which we build this measure is always a multiple of 10ms. In our baseline specifications we set the interval to be exactly 10ms, but do some robustness analysis using longer windows.[8] The fact that our order book data is on a 10ms sampling frequency and trades are sampled more frequently also means that there will be some noise in our GL measure. Assume that we are measuring GL over precisely a 10ms interval. A trade arriving just after an order book snapshot will see the majority of this 10ms interval coming after the trade, while a trade arriving just before an order book update will have most of the 10ms interval pre-trade. Thus, while in this example depth changes are always measured over a 10ms interval, there will be small variations across trades in the portion of that interval that comes before the trade and the portion that comes afterwards.

In the definition above, depth measures $PREQTY_{qv}^{ask}(.)$ and $POSTQTY_{qv}^{ask}(.)$ are quantities available in the order book of venue *qv* within a certain distance of the midquote. To measure this distance, we look at the distribution of the difference between third most competitively priced limit buy and sell orders from the consolidated order book and take the 90th percentile of that distribution. This 90th percentile is used to define a stock-specific band around the current midquote such that only orders within that band contribute to the GL measure. We have chosen its width to ensure that we capture a decent amount of order activity, while excluding orders that lie a long way from the stock's midquote. This focuses attention on cancellations of those orders with prices close to the execution price on *tv* and thus which are most likely to be relevant to GL measurement.

The baseline GL measure above is trader, trade time, stock, venue, and side specific, and we want to aggregate these data to that they can be compared across stocks and times. To make the

[8] Other time intervals considered are 20ms, 50ms, and 100ms. There are all below human reaction time.

data comparable across stocks, and to aggregate up to the daily level we express GL as a proportion of displayed quantity. First, we compute the following measure:

$$GL_{tv\to qv}(\Delta\tau;i;d;m)=\frac{\sum_{\tau\in d} GL^{bid}_{tv\to qv}(\tau;\Delta\tau;i;m)+\sum_{\tau\in d} GL^{ask}_{tv\to qv}(\tau;\Delta\tau;i;m)}{\sum_{\tau\in d} PREQTY^{bid}_{qv}(\tau;i;m)+\sum_{\tau\in d} PREQTY^{ask}_{qv}(\tau;i;m)}. \quad (2)$$

In Equation (2), trade-time measures $GL_{tv\to qv}(.)$ and $PREQTY_{qv}(.)$ are summed for all trades within a given day to give aggregated GL for member *m* on venue *qv* in response to executions on venue *tv* on day *d* for stock *i*.

Next, we construct a weighted average GL across members, where the weight for member *m* is equal to the average contribution of that member to the depth of stock *i* on the quote venue over the day considered. Finally, monthly averages of daily mean GL are computed for each stock.

Equation (2) gives the GL supplied by a member as a fraction of the total depth attributable to that member on the quote venue. When aggregated up, this gives a sense of the fraction of liquidity supplied to that venue that is likely to disappear as a result of a trade on another venue. An alternative way to scale GL is to divide it by the size of the original trade on venue *tv*. This allows us to ask, for example, if a trade on one venue leads to the removal of a similarly sized order on another venue.

Thus, we construct an alternative GL measure where, in the denominator of the computation, we replace the pre-trade depth contributed by member *m* on venue *qv* with the size of the trade that triggered the GL measurement. In our empirical work, we perform all of our estimations using both GL measures that scale by depth and using GL measures that scale by trade size.

In our summary statistics we present cross-stock averages of GL per pair of venues. For each pair of venues, the average computed reflects the mean level of GL on the quote venue (*qv*) observed due to executions on the trade venue (*tv*). We also wish to compute a single number to summarize the scale of the GL problem on a single venue. This entails averaging across trade venues to focus on a single quote venue. The weight used in this averaging for venue *tv* is equal to the total volume executed on *tv* over the sample divided by the sum of the volumes on all three trade venues.

5.2. Measuring order book refilling in the next 10ms after GL cancellations

One may argue that our GL measure is not necessarily capturing ghost liquidity posted to optimize execution probabilities but that it could reflect quote updates in reaction to information contained in trades on other venues. If these quote updates are due to orders being re-priced, we should observe order cancellations and then swift resubmissions at different prices but for roughly the same quantity in the GL venue's order book. No such resubmissions should occur in the case of genuine GL. Thus, in order to distinguish GL from quote updating, we compute a book refill rate for the 10ms after the time window over which GL is measured. For a given member whose order cancellation has contributed to our GL calculation, this refill rate equals the liquidity added by that same member on the same venue where GL is being measured.[9] To be more explicit about the calculation of the refill rate, let us return to the example we used when discussing the GL calculation in Equation (1). At time *t*, a limit sell order submitted by member *m* is executed on venue *tv* for stock *i*. At the same time, *m* also has limit sell orders posted on venue *qv* for stock *i*. We measure the sell-side GL of *m* on venue *qv* by looking at her cancellations inside a 10ms time window that starts at the closest 10ms timestamp preceding trade time τ. The refill rate is calculated over the next 10ms window in the following way:

$$Refill^{ask}_{tv\to qv}(\tau+10ms;i;m)=$$

$$\left(POSTQTY^{ask}_{qv}(\tau+10ms;i;m)-PREQTY^{ask}_{qv}(\tau+10ms;i;m)+\sum_{\tau+10ms}Volume^{buy}_{qv}(i;m)\right)\Big/GL_{tv\to qv}(\tau;10ms;i;m) \quad (3)$$

where $PREQTY^{ask}_{qv}(\tau+10ms;i;m)$ is the total limit sell order quantity posted by trader *m* on venue *qv* at the first 10ms order book snapshot following trade time τ (on venue *tv)* and $POSTQTY^{ask}_{qv}(\tau+10ms;i;m)$ is the total limit sell order quantity posted by member *m* on that same venue 10ms later. $\sum_{\tau+10ms}Volume^{buy}_{qv}(i;m)$ consists of all executions against trader *m*'s limit sell orders on *qv* in that same 10ms window starting after the initial trade. When added to the difference in quantities, it yields the amount of liquidity that member *m* has added to the quote

[9] Order submissions are only counted towards the refill quantity if they are submitted within a certain distance of the midquote. This distance is the same as that defined above for the GL computation and the midquote we use is that observed at the end of the GL measurement window.

venue book immediately after the GL cancellations. This is then expressed as a percentage of the 10ms GL measured for the same trade and the same member *m* on the quote venue. A positive refill rate indicates that members refill the book after cancelling orders whereas a negative refill rate indicates that the members continued cancelling liquidity after the end of the GL window. Those refill rates are computed for all trades which generated positive GL and are then averaged across time, members, and stocks, by countries, platforms, stock terciles, and member categories.

5.3. Descriptive statistics for GL

We present several descriptive statistics in order to understand how GL is distributed geographically and whether there is any relationship with market size. We also analyze whether GL is different across member categories.

Table 3 about here

Panel A of Table 3 reports GL by country and is obtained by averaging across primary exchange and alternative venues. This panel reveals some country-level heterogeneity with GL varying between 0.23% and almost 7%. The countries with the highest GL are the Netherlands, the UK, and Belgium whereas Spain and Italy exhibit much lower GL. Panel A also indicates that the average level of GL for each country does not change much as one moves from a 10ms GL measurement window to a 100ms window. Finally, Panel A also shows that the refill variable is, on average, close to zero for all countries. This suggests that our GL measure is not contaminated by cancellations due to members repricing orders in response to trades on other venues.

Panel B reports GL by platform, by taking the weighted average across the trades that trigger our measurements. We find that GL is much smaller on primary exchanges in comparison with the three alternative venues.

Panel C breaks down GL by pairs of venues. The first column of the table gives the name of the venue where GL is being measured and the second column gives the name of the venue where the trade that triggers the measurement occurred. For example, a trade on Chi-X leads to a 3.74% reduction in outstanding limit orders by that same member on the primary exchange, on average. The results show that the proportion of limit order volume that is removed by the same member on another platform ranges from roughly 2% to almost 9%. The results also reveal that there are no big differences across trade venue-GL venue pairs. The small differences also seem to be

unrelated to the type of venue (i.e., alternative-primary exchange or alternative-alternative) pairs. As in Panel A, the average value of the refill rate is always close to zero.

Table 4 about here

Table 4 has the same structure as Table 3, except it reports figures based on GL as a fraction of trade size rather than quantity outstanding. This has the effect of greatly increasing the mean value of the GL variables to 19% on average across all stocks, and to more than 30% in some cases. Thus, for example, after a trade on a UK venue, one subsequently sees around 40% of the trade quantity cancelled on a different venue by the same trader. The UK and the Netherlands are still among the countries with the highest GL and Spain is still the lowest. The difference across venues in average GL as a fraction of trade size is now fairly small with, if anything GL being larger on the primary exchange. Looking at pairwise average GL levels, it is clear that GL on alternative venues when a trade occurs on the primary tends to be much smaller than GL on alternative venues when the triggering trade is on a different alternative venue.

We have also computed the same measures as in Tables 3 and 4, but for a modified GL measure. The GL metric that we have worked with thus far, i.e., equation (1), subtracts the aggregate quantity traded in the interval from the difference between pre and post-trade liquidity outstanding, so as not to include involuntary reductions in liquidity associated with trades in the GL measure. However, some of these trades may have been executions of genuine ghost orders by counterparties with fast, smart-order routing technology (i.e., by agents whose technology is fast enough to allow them to hit duplicate orders on multiple venues before the liquidity suppliers can remove them). Thus, our GL measure represents a lower bound on true ghost liquidity. To provide an upper bound, we also compute summary statistics for a GL measure which is just the change in liquidity pre-trade to post-trade. This modified measure implicitly assumes that all executions against this member and in this stock in the interval were of ghost orders. The adjustment roughly doubles the level of GL measured as a fraction of quantity from just over 4% (measured across all stocks) to almost 9%. On some markets and some venues, GL reaches 15%. Thus, allowing executed volume to be thought of as GL significantly increases the scale of GL. Performing the

same adjustment to our GL measure based on trade size leads to statistics in which GL rises from roughly 20% to 25%. Thus, there is a rise here too, but proportionately less big.[10]

Table 5 about here

Returning to the original GL measure, we proceed to investigate the variation of GL with stocks' activity levels. Table 5 displays the average level of GL per market value tercile. Differences in GL expressed as a fraction of pre-trade liquidity are in general not very large, but there is a tendency for GL to rise with market cap. This tendency is much more clear when GL is expressed related to the size of the triggering trade. The table also demonstrates that GL is negatively related to volatility in a stock, presumably because the costs of order duplication across venues (e.g., multiple executions and thus over-trading) are larger in a more volatile world. Finally, as one would expect, the last panel of Table 4 shows us that GL is larger on average in stocks with more fragmented trading.

ESMA (2016) find that the cross-stock covariances of order duplication intensity with market cap, volatility and fragmentation have the same sign as the covariances between our GL measure and those variables. They also find that the likelihood of duplicate orders being cancelled also tends to rise with market cap and fragmentation. Thus, their results and ours are consistent.

Table 6 about here

It is important to understand whether GL is mainly due to some categories of members. Table 6 decomposes average GL by members according to their trading scope (*local trader* and *global trader*) and trading aggressiveness (*liquidity taker* and *liquidity supplier*). We further distinguish according to their trading speed (*Slow, AT* and *HFT*) and their capacity (*Agent* or *Principal*). The most interesting differences arise when comparing members acting as principal and those acting for their clients and when comparing traders by speed. As we would expect, the average GL for HFTs is, at 5.75% of their total pre-trade liquidity, about 1.5 times larger than the average GL associated with algo traders (AT) which is, in turn, around 1.4 times larger than GL from slow traders. Thus, HFT trading strategies lead to greater duplicated liquidity. ESMA (2016) report a similar finding for their direct analysis of order duplication. GL is also typically higher when

[10] Tables of summary statistics for the adjusted GL measures, identical in structure to Tables 3 and 4, are available on request from the authors.

members are acting as principal rather than agent. This feature is strengthened by the fact that members acting as agent have the greatest refill rate (3.16%).

Let us recall that the starting point of a GL calculation is a trade on a given venue. At the time of the trade, the passive counterparty may or may not have duplicated limit orders on the venue where GL is measured. For that reason, we also provide, in Table 6, the percentage of trades for which there is order duplication on the GL venue. By definition, this percentage is extremely low for local traders (3.31%), but in those few cases where they duplicate orders, the average value of their GL is more than half of that of global traders. Another striking case is that of members trading as agent. They duplicate limit orders far less often than members trading as principal (16.78% vs. 51.23%), but when they do so, their level of GL reaches one half of that of members trading as principal.

The fact that on average GL differs systematically across member categories suggests that it may be important to control for such categories in our multivariate analysis. We now turn to our empirical model and identification strategy.

6. Determinants of Ghost Liquidity

In this section we set out to identify the determinants of GL. Before doing so, we first develop a set of hypotheses underpinning our empirical work. Our analysis is based on the idea that, when the order flow in a stock is fragmented across several order books, limit order traders may increase their expected liquidity-providing profits by posting GL. Duplicating liquidity supply across books, with the intention to cancel residual orders as soon as the desired quantity is executed in one book, increases expected market-making profits by reducing both execution delays and non-execution risk. Yet this improvement in execution speed and probability of execution is effective if marketable orders actually arrive on several venues, i.e., if the order flow is fragmented enough. We thus expect GL to increase with fragmentation (Hypothesis H1). Also, the incentive to post GL is greater when other options to improve execution probability, such as competing on price, are not available. GL should then be greater when the tick size is more likely to be a binding constraint on price competition (i.e., when a large tick size makes price undercutting expensive or impossible). For that reason, we expect GL to increase in the tick size (Hypothesis H2). By definition, GL is a tool used to increase the profits of limit order traders when making markets. We thus expect frequent liquidity suppliers (Hypothesis H3) and traders acting as principal

(Hypothesis H4) to post more GL than otherwise similar traders. The eagerness of a liquidity supplier to trade depends on her pre-trade inventory level. An inventory that strongly deviates from its optimal level gives a greater incentive to seek execution speed by posting GL. From there, we hypothesize that the GL posted by a market member increases with the deviation of her stock inventory from normal level (Hypothesis H5).

However, the potential benefit of GL for a limit order trader comes at the cost of the risk of over-trading, i.e., the risk of being executed at multiple locations such that total quantity traded exceeds desired quantity. Any factor impacting this risk is also a potential determinant of GL. As over-trading risk is realized when duplicated orders are hit before being cancelled, the trading speed of the GL trader relative to others is obviously crucial. We expect the GL of a market member to increase with her trading speed advantage (Hypothesis H6). Her trading speed advantage also depends on the technology used by those she is trading against. In particular, the trading speed advantage she uses for fast cancellations will not be effective if, on the other side of the market, sophisticated market order traders use smart order routers (SORs) to hit her limit orders on several platforms simultaneously. We thus posit that GL decreases with the presence of SORs (Hypothesis H7). Finally, trading speed advantages are better exploited on platforms with lower latency. This leads us to expect GL to be greater on alternative platforms (Hypothesis H8).

We test these eight hypotheses by conducting a panel regression analysis of data measuring the GL of global members on a set of control variables. We aggregate data to a 15-minute sampling frequency before running the regressions. We then refine the analysis by analyzing data for specific sub-populations of the set of global members. We finish by providing evidence that GL is not the result of shifts in liquidity by the same member from the GL venue towards the trading venue. We do so by computing the added liquidity on the trade venue as well as a GL consolidated across platforms.

6.1 Global members

The left-hand side variable in our regression analysis is the stock- time- and member-specific GL measure defined by Equation (2) and in our base model Δt = 10ms. As mentioned above, for this analysis we have aggregated GL to a 15-minute sampling frequency.

Our regression model is

$$
\begin{aligned}
&GL_{tv\to qv}\left(\Delta t;i;t;m\right)\\
&=\alpha+\beta_1 FRAG_{i,t-1}+\beta_2 TICK_{i,t}+\beta_3 AGENT_{i,m}+\beta_4 LS_{i,m}+\beta_5 INV_{i,t-1,m}\\
&\quad+\beta_6 HFT_{i,m}+\beta_7 AT_{i,m}+\beta_8 SOR_{i,t-1}+\beta_9 SOR_{i,t-1}^2+\beta_{10} PEtoALT_{tv,qv}+\beta_{11} ALTtoPE_{tv,qv}\\
&\quad+\gamma_1 GL_{HFT\backslash i,t,m}^{Others}+\gamma_2 GL_{AT\backslash i,t,m}^{Others}+\gamma_3 GL_{Slow\backslash i,t,m}^{Others}\\
&\quad+\delta_1 VOLUME_{i,t}+\delta_2\sigma_{i,t-1}+\delta_3 PRICE_{i,t}+\delta_4 TRADESIZE_{i,t,m}\\
&\quad+\mu_1 IMB_{i,t}+\mu_2 IMB_{i,t-1}+\varepsilon_{i,t,m,tv,qv}.
\end{aligned}
\tag{4}
$$

$GL_{tv\to qv}\left(\Delta t;i;t;m\right)$ is the aggregated GL on venue *qv* resulting from a trade on venue *tv*, for stock *i*, on 15-minute period *t*, and for member-capacity *m*. Suggested by our hypotheses above, our key explanatory variables of interest are the fragmentation level in the stock, the tick size, the member characteristics, the presence of SORs, and the platforms' characteristics. We control for the GL of other members, the usual determinants of liquidity including volume, volatility, and price level, as well as some order flow characteristics, namely trade imbalance and trade size. We further include stock -fixed effects and intraday time fixed effects identifying each 15-minutes period of the trading session.

$FRAG_{i,t}$, the degree of fragmentation of stock *i* in period *t*, is the reciprocal of a Herfindahl-Hirschman index based on the market shares in volume of the four trading platforms.[11] $TICK_{i,t}$ is the tick size of stock *i* divided by the closing price of the day.

The market member characteristics consist of four dummy variables, $HFT_{i,m}$, $AT_{i,m}$, $AGENT_{i,m}$, and $LS_{i,m}$, identifying the trader type, plus a variable measuring the level of the member's stock inventory in the previous period, denoted $INV_{i,t-1,m}$. Dummies $HFT_{i,m}$, $AT_{i,m}$, $AGENT_{i,m}$, and $LS_{i,m}$ are equal to one when in that stock, a market member is an HFT, an AT, trading as agent, or a liquidity supplier respectively, and zero otherwise. The inventory variable $INV_{i,t-1,m}$ is the absolute value of the member's inventory over the preceding 15 minutes. As in Hansch, Naik and Viswanathan (1998), we compute member *m*'s inventory in stock *i* in each interval, then standardize (by subtracting by the mean level of the inventory for member *m* and stock *i* and scaling by the standard deviation of that member's inventory in that stock) and, finally, take the

[11] This type of measure is commonly used in the literature on market fragmentation (see Degryse et al., (2015) and Gresse (2017)). In terms of interpretation, our *FRAG* index ranges from one to four, one indicating no fragmentation, or in other words, a consolidation of volumes on a single venue, and four indicating maximum fragmentation, that is volumes equally distributed across the four venues. A *FRAG* index of two would mean that the level of fragmentation is equivalent to the maximum level of fragmentation between two markets, i.e., 50% of the volumes on each.

absolute value. The measure thus represents the distance between current inventory and its 'normal' level for that member and stock. If the simultaneous submission of orders to multiple venues is used by traders to manage extreme inventories towards zero, then we might expect a positive relationship between our inventory variable and GL.

$SOR_{i,t}$ is a proxy for the intensity with which smart order routing algorithms are being employed in the trading of stock *i* in period *t*. We judge a member to be using smart order routing when she is engaging in aggressive trading in the same stock on multiple venues simultaneously. By aggressive trading, we mean trading generated by market orders or marketable limit orders at prices within the bid-ask spread used to measure GL (cf. Section 5.1). For stock *i*, member *m* and a particular pair of trading venues, we compute the quantity simultaneously aggressively bought in an interval of 10ms as;

$$SOR(i;m;buy) = 2 \times \min\left[volumebuy(i:m;A), volumebuy(i:m;B) \right] \quad (5)$$

where *A* and *B* are the two trading venues. We compute a similar quantity for sell volumes. We then aggregate across members and the buy and sell sides of the market to give aggregate smart-order routing trading in stock *i* for the chosen interval of time and for the pair of venues *A* and *B* and, finally, we scale this measure by total buy and sell volume in the stock in the interval. We expect an increase in smart order routing to be associated with a decrease in the supply of ghost liquidity to venues, as the risk of multiple executions and thus over-filling is increased.

Both the *FRAG* and the *SOR* variables are introduced with a lag in the regression so as to consider causal effects rather than correlation effects.

The platform characteristics capture whether *tv* and *qv* are the primary exchange (*PE*) or one of the alternative venues (*ALT*). $PEtoALT_{tv,qv}$ is one when trade venue *tv* is *PE* and the venue on which we measure GL (i.e., *qv*) is *ALT*, zero otherwise. $ALTtoPE_{tv,qv}$ has a similar interpretation. The base case is where *tv* and *qv* are both *ALT*.

We further control for the GL by other HFT members ($GL^{Others}_{HFT\backslash i,t,m}$), other AT members ($GL^{Others}_{AT\backslash i,t,m}$), and other slow traders ($GL^{Others}_{Slow\backslash i,t,m}$) excluding member *m* (denoted by *m*) in period *t* for stock *i*. We also include a set of stock-time characteristics known as liquidity determinants. Volatility $\sigma_{i,t}$ is a price range computed as the difference between the highest and the lowest prices of stock *i* over 15-minute interval *t* scaled by the middle. $VOLUME_{i,t}$ is the logarithm of the total euro volume traded in stock *i* on the four venues over period *t*. $PRICE_{i,t}$ is the last cross-

venue log midquote on the day of period *t* for stock *i*. As a trade characteristic we include the trade size on *tv* when GL on *qv* is scaled by displayed quantities. This variable, denoted $TRADESIZE_{i,t,m}$, equals the average size of the trades executed on *tv* and triggering GL measuring on *qv* for member *m*, stock *i*, and period *t*. Size is measured as the log of the euro value of the trade. The variable is abandoned when GL is already scaled by trade size.

Finally, we control for past and contemporaneous order imbalance, respectively denoted $IMB_{i,t-1}$ and $IMB_{i,t}$, to make sure that GL is not driven by trade-conveyed informational effects. $IMB_{i,t}$ is the absolute value of the difference between aggressive buy and sell trading volumes, expressed as a percentage of the total traded volume on all platforms for stock *i* in period *t*.

Table 7 about here

The first four columns of Table 7 display the results for our empirical model using "GL as a percentage of pre-trade liquidity" employing different time windows ranging from $\Delta\tau = 10$ms in the first column, to $\Delta\tau = 20$ms, 50ms and 100ms in the second, third and fourth columns, respectively. We employ a Tobit model as our dependent variable has truncations at zero and one, i.e., in many instances there is no withdrawal of liquidity (GL=0), or all liquidity is withdrawn (GL=1). The last column in Table 7 presents the results where we scale GL by the trade size at the trading venue *tv*. Here we use a Tobit model with truncations at zero.

We first examine the impact of member characteristics – our key variables of interest. Consistent with H6, all columns of Table 7 show that trades where limit orders posted by fast traders (both HFTs and ATs) are executed lead to significantly more GL than otherwise similar trades against slow traders (the base case) and that HFTs post more GL than ATs, with a statistical significance at the 1% level. In particular, based on the first column ($\Delta t = 10$ms), an HFT (AT) member withdraws 7.88 (2.80) percentage points more of its outstanding limit orders on venue *qv* following the execution of one of its limit orders on venue *tv* compared with a slow member in a similar situation. HFT members thus post just over five percentage points more GL than AT members. GL as a percentage of pre-trade liquidity is also more pronounced when a member (i) behaves as a liquidity supplier (2.58 percentage points), and (ii) acts as principal (2.03 percentage points, i.e., *AGENT*=0). Results for longer time windows displayed in the second to fourth column are comparable.

The standardized, absolute inventory variable has a significant and negative coefficient in our regressions. The more extreme inventory positions are associated with smaller GL. This suggests that members do not use GL to manage inventory in times when inventory is extreme. Instead, the sign of the coefficient is consistent with members building up inventories using GL strategies. This suggests that we should reject our hypothesis H5. Having said this, the economic magnitude of the coefficients is small, with a one standard deviation increase in inventory leading to a fall in GL of around 0.1 percentage points.

The last column in Table 7 presents the results where we scale GL by the trade size at the trading venue *tv*. It allows us to assess what fraction of the trade size executed on the trading venue *tv* is withdrawn by members on the quoting venue *qv*. HFT members on average withdraw 22 percentage points more of the trade size compared with slow members, and around 15 percentage points more when compared with AT members. AT members withdraw on average 5.5 percentage points more than slow traders. This is again consistent with H6. Members acting as agent and liquidity suppliers withdraw 5 percentage points less and 8.5 percentage points more than principal traders and liquidity takers respectively, consistent with hypotheses H3 and H4. All of these effects are significant at the 1% level.

We now turn to all other characteristics and focus on the results presented in columns 1 to 4. The row on "trade characteristics" shows that larger trades are associated with greater GL. Members have more incentives to cancel orders when trade size on the trading venue is larger. Results for $\Delta t = 10$ms (first column) show that when trade size doubles, GL increases by 1.2 percentage points.

The next rows in Table 7 show the results for the "platform characteristics". Based on column 1 ($\Delta\tau = 10$ms), the *PEtoALT* coefficient shows that GL is 1.8 percentage points less pronounced when the trade takes place on the primary exchange and the GL venue is another venue compared with the base case *ALTtoALT*. The coefficient on *ALTtoPE* is significant, positive and larger in magnitude than that on *PEtoALT* across columns (1)-(4). In sum, GL is least pronounced when trades take place on the primary exchange and most pronounced for trades occurring on alternative venues and where the liquidity is then cancelled on the primary exchange, in line with H8.

Our regression model controls for other member groups' GL activity on that day for that stock. In general, we find that a member's GL seems to co-move with the GL of other members. This effect is most pronounced when other HFTs and ATs are active posters of GL.

Next, we discuss the results for the impacts of order flow and stock characteristics. Across the various time windows, the significant positive coefficients on trading volume and fragmentation imply that GL is greater for stocks that are traded more heavily and on a dispersed set of platforms (in line with H1). Absolute order imbalance has a consistent and significant negative effect. We were concerned that the cancellation activity behind GL might be generated by members revising stock valuations due to the information contained in trades. Neither past order imbalance nor contemporaneous order imbalance positively impacts GL, which is not in line with an information-based interpretation. GL significantly increases with an increase in the price range for stock *i* and is smaller for stocks with larger tick sizes. The second result is inconsistent with our hypothesis H2, which suggested that GL might be more intensively used when undercutting by price is more difficult.

Finally, there is a concave relationship between smart order routing and GL. This generates small increases in GL when smart order routing is scarce but rising, but very large negative effects when smart order routing is large and rising (e.g. if smart order routers were only 20% of the trade population, GL would be 4 percentage points greater than if SOR was zero, while if SOR was at 80% of trading, GL would be almost 25 percentage points lower). So, when smart order routers are used extensively, we see low use of GL, likely due to the multiple execution risk that SOR technology exposes the users of GL to (i.e., in line with H7)

6.2 Member categories

Table 8 shows the results of Equation (4), where Δt = 10ms, for subsamples that focus on various member categories. This allows us to study whether particular determinants are more relevant for some member categories: column (1) focuses on all members that are "fast traders"; columns (2) and (3) split up these fast traders into ATs and HFTs; Columns (4) and (5) display results for "Liquidity suppliers" and "Fast liquidity suppliers".

Table 8 about here

The coefficient on *HFT* in column (1) shows that HFTs withdraw 5.51 percentage points more of their pre-trade liquidity on the quoting venue than ATs (i.e., the base case) following a trade on the trading venue. Compared with the first column of Table 7 presenting results for all member categories, some interesting differences in the magnitudes of our control variables can be observed. First, the positive coefficient on *ALTtoPE* in the regressions in Table 7 appears to be driven by the

behavior of ATs, with this coefficient being negative and around the same magnitude as *PEtoALT* for HFTs.

Next, co-movement of GL is most pronounced among own-member types. Columns (2) and (3) for example show coefficients of around 0.15 for GL_{AT}^{Others} and 0.20 for GL_{HFT}^{Others}, respectively. These are considerably larger than the coefficients on other member categories.

The coefficients on order flow characteristics and on trader inventory are consistent in sign with those in Table 7 while, within the stock characteristics, volatility again has a positive and significant effect in the main. The coefficient on tick size is significantly negative in 2 of the 5 regressions, namely those corresponding to liquidity suppliers.

All effects mentioned above are statistically significant at the 1% level.

6.3 Alternative explanations: Is ghost really ghost?

In this subsection, we discuss and rule out possible alternative explanations. One possibility is that members move their orders from the GL-venue to the "venue where the action takes place", i.e., the trading venue, in order to increase their execution probability. In that event, what we call ghost liquidity would simply reflect a reshuffling of liquidity towards the trading venue.

To study this alternative explanation, we first check whether orders cancelled on the quote venue (GL) are swiftly resubmitted on the trade venue in the same and the next 10ms windows. According to our observations this is not the case. On average, across all stocks, 15.6% of the GL measured on the quote venue is also cancelled by the same member on the trade venue and refill rates on the trade venue in the next 10ms are close to zero.

Second, to dig further, we also take an aggregate perspective and focus on the evolution of a member's *consolidated depth across all venues* around a trade. In particular, we study how a member's offering of market depth across all venues (i.e., all its outstanding limit orders on the side of the trade in all venues (trading and ghost venues)) evolves in the time window before (i.e, at t) to after (i.e., to t+10ms) the trade taking place on a trading venue. We again scale this difference in depth either by a member's pre-event consolidated depth, or by the size of the trade, and control for trades against our member in the event window. We find that on average it equals 6.62% of a member's consolidated depth and 59.09% of trade size. Since these numbers are larger than our cross-venue liquidity measures (4.04% and 19.67%, respectively), we find that a member is not shifting its limit orders to the trading venue. In contrast, a member further seems to withdraw liquidity also at the trading venue. In addition, we follow subsection 5.2 and study whether orders

that are cancelled in the consolidated order book are not refilled within the 10ms following the time window over which GL is measured (i.e., the "refill rate"). On average, we find a negative refill rate of -2.84% of the globally cancelled liquidity of that member, indicating that members continued cancelling liquidity in the next 10ms.

7. Impact of Ghost Liquidity on trading costs

Finally, we analyze how the use of ghost liquidity strategies affects the trading costs of various trader groups. We might expect markets with greater incidences of GL to be those in which 'genuine' liquidity is harder to measure and so execution cost management might be less effective. This may mean that GL is positively correlated with costs of trading.

We test this hypothesis by running panel regressions of daily effective spreads by stock and venue on various conditioning variables, including the GL measure for that stock, venue and day and the product of the GL measure and a primary exchange dummy. We compute daily GL for a particular venue by taking the measures computed earlier for pairs of trade venues and quote venues, fixing a particular quote venue and aggregating across trade venues. The specification is as follows:

$$ES_{i,d,k} = a + b\ RVolat_{i,d} + cVOLUME_{i,d} + d\ PRICE_{i,d} + eTRADESIZE_{i,d,k} + fPE_{i,k} + gGL_{i,d,k} + hGL_{i,d,k} \times PE_{i,k} + iES_{i,d-1,k} + \varepsilon_{i,d,k}. \quad (6)$$

In this equation, $ES_{i,d,k}$ is the average effective spread for stock *i* on day *d* on venue *k*. $RVolat_{i,d}$ is the realized volatility in that stock, computed as the square root of the average squared five-minute logarithmic return of stock *i* on day *d*. $VOLUME_{i,d}$ is the logarithm of the total euro volume traded in stock *i* on all four venues on day *d*. $TRADESIZE_{i,d,k}$ equals the average size of the trades that were used to construct the effective spread variable. $PE_{i,k}$ is a dummy equal to one when the venue for which we are computing effective spreads is the primary exchange, zero otherwise. $GL_{i,d,k}$, which is the ghost liquidity measured on venue *k* for stock *i* on day *d*, is the variable of interest. We examine whether the impact of ghost liquidity on effective spreads may differ between the primary exchange and alternative venues by interacting $GL_{i,d,k}$ with the primary exchange dummy. Autocorrelation is accounted for by including the first lag of the dependent variable, and stock- and day-fixed effects are included.

We run three versions of regression (6), the difference between them being the specification of the dependent variable. In the three regressions it is measured as the effective spread paid by slow liquidity takers, algorithmic liquidity takers and HFT liquidity takers, respectively.

Table 9 contains estimates of this model. There are several familiar results in the table (e.g., spreads increase with volatility, decrease with volume, increase with trade size and are positively autocorrelated). As for the coefficients on GL, they are all positive and two are significant. These are GL in the algo trading regression and the interaction of GL and the primary venue dummy in the slow trader regression. Thus, algo liquidity takers pay more when GL is large, whatever the venue under consideration, while slow traders pay more when GL is large on the primary venue. The effect on slow traders being focused on the primary venue makes sense as this is where they likely do the vast majority of their trading. HFTs do not suffer from GL at all (presumably because they are sophisticated enough to evaluate its effects).

Tables 9 about here

Table 9 also shows the results where we replace 'GL' and 'GL×primary exchange' by 'GL of HFTs' and 'GL of HFTs×primary exchange' as our main explanatory variables. We ask the question whether GL stemming from HFTs influences effective spreads for the various trading groups differently than overall GL. We find that only slow traders on the PE face somewhat higher effective spreads with the GL of HFTs, and the economic magnitude is somewhat larger compared to the impact of overall GL. Other trader groups seem not affected by the GL of HFTs.

8. Conclusion

The objective of this paper is to assess the scale of Ghost Liquidity (GL) and the factors that drive it in fragmented markets. GL is related to limit order duplication across venues. We define it to exist when, in response to the execution of a limit order on a particular venue, the submitter of that order swiftly cancels similar orders on other venues. Such liquidity provision strategies are built to maximize execution probabilities. On the one hand, they may benefit cross-market liquidity by improving execution probabilities, yet on the other hand, GL may mislead market participants in their perception of the true liquidity available in the marketplace.

By drawing on a unique data set that covers the primary exchange and the three main alternative trading venues in Europe, i.e., Chi-X, BATS, and Turquoise, for 91 European stocks primary listed

in nine countries, we find that GL is an economically significant phenomenon that deserves attention from market participants and regulators. Limit order duplication is however not always GL. In the presence of duplicated limit orders, for 100 shares traded on one venue, the submitter of the passive order removes on average around 20 shares from the order book of another venue. At the market level, over 4% of the consolidated depth is GL, this average percentage being greater on alternative venues (between 6% and 7%) than on primary exchanges (3.43%). Those figures are not sizeable enough either to challenge the depth improvement related to fragmentation found by Degryse et al. (2015) and Gresse (2017), or to create severe instability in total liquidity. Furthermore, GL does not necessarily affect all traders in the same way, as fast traders using properly calibrated smart order routers may catch GL before it is withdrawn.

GL may however reach substantial levels for some stocks, platforms, or traders. The cross-section of our sample shows that GL is greater for larger, more fragmented stocks and less volatile stocks. Further, GL increases with trading volumes, trade size, and market fragmentation. It decreases when smart order routing is particularly prevalent. HFTs, traders acting as principal, and traders implementing multi-market market-making strategies post more GL than others. Further, regarding HFTs, their use of GL is the highest when they duplicate limit orders across alternative platforms. Those results are robust to changes in the time window used to measure GL, and they are not significantly impacted by cancellations due to quote updating in response to trades.

In our final piece of analysis, we find that GL causes the execution costs of algorithmic traders to rise and GL on the primary venue causes the execution costs of slow traders to rise. Thus, there is evidence that this phenomenon disrupts the execution cost management strategies used by all traders aside from HFTs.

Overall, we show that ghost liquidity is a significant phenomenon in European equity markets, and it has direct impact on the trading costs of those executing in those markets. A consequence of our findings is that simple consolidated liquidity measures may overestimate true liquidity in fragmented electronic markets. On the flip-side, previous research shows that fragmented markets tend to be, on average, more liquid than consolidated ones and our estimates suggest that, while at particular times and for particular stocks, GL is large, on average it is not large enough to outweigh the positive effects of fragmentation on market liquidity.

References

Blocher, Jesse, Cooper, Rick A., Seddon, Jonathan, and Van Vliet, Ben (2016). Phantom Liquidity and High Frequency Quoting. *Journal of Trading*, 11(3), 6-15.

Bouveret, Antoine, Guillaumie Cyrille, Aparicio Roqueiro, Carlos, Winkler, Christian, and Nauhaus, Steffen (2014). High-frequency trading activity in EU equity markets. ESMA Economic Report #2.

Brogaard, Jonathan (2010). High frequency trading and its impact on market quality. Northwestern University Kellogg School of Management Working Paper, 66.

Brogaard, Jonathan, Hendershott, Terrence J., and Riordan, Ryan (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8), 2267-2306.

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

Chen, Haoming, Foley, Sean, Goldstein, Michael, and Ruf, Thomas (2017). The Value of a Millisecond: Harnessing Information in Fast, Fragmented Markets. Working Paper available at SSRN: https://ssrn.com/abstract=2860359 or http://dx.doi.org/10.2139/ssrn.2860359.

Dahlström, Petter, Hägströmer, Björn, and Nordén, Lars L. (2018), Determinants of Order Cancellations, available at SSRN: https://ssrn.com/abstract=3012831

Degryse, Hans, de Jong, Frank, and van Kervel, Vincent (2015). The impact of dark and visible fragmentation on market quality. *Review of Finance*, 19(4), 1587-1622.

ESMA (2016). Order duplication and liquidity measurement in EU equity markets. ESMA Economic Report No. 1, 2016.

Foucault, Thierry, and Menkveld, Albert (2008). Competition for order flow and smart order routing systems", *Journal of Finance*, 63(1), 119-158.

Gresse, Carole (2017). Effects of lit and dark market fragmentation on liquidity. *Journal of Financial Markets*, 35, 1-20.

Hansch, Oliver, Naik, Narayan Y. and Viswanathan, S. (1998). Do inventories matter for dealership markets? Evidence from the London Stock Exchange. *Journal of Finance*, 53(5), 1623-1656.

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

Hasbrouck, Joël, and Saar, Gideon (2013). Low-latency trading. *Journal of Financial Markets*, 16(4), 646-679.

Hendershott, Terrence J., Jones, Charles J., and Menkveld, Albert J. (2011). Does algorithmic trading improve liquidity?. *Journal of Finance*, 66(1), 1-33.

Jovanovic, Boyan and Menkveld, Albert J. (2016). Middlemen in limit order markets. Working Paper available at http://dx.doi.org/10.2139/ssrn.1624329.

Kirilenko, Andrei A., Kyle, Albert S., Samadi, Mehrdad, and Tuzun, Tugkan (2017). The flash crash: High frequency trading in an electronic market, *Journal of Finance*, 72(3), 967-998.

Menkveld, Albert J. (2013). High frequency trading and the new market-makers", *Journal of Financial Markets*, 16(4), 712-740.

Menkveld, Albert J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics*, 8, 1-24.

O'Hara, Maureen, and Ye, Mao (2011). Is fragmentation harming market quality?", *Journal of Financial Economics*, 100(3), 459-474.

Yueshen, Bart Z. (2014). Queuing uncertainty in limit order market. Insead Working Paper.

van Kervel, Vincent (2015). Competition for order flow with fast and slow traders. *Review of Financial Studies*, 28(7), 2094-2127.

Table 1. Descriptive statistics on sampled stocks

Country	Number of stocks		Market value (EUR Mn)	Value traded (EUR Mn)	Cross-market bid-ask spread	Market share of the primary exchange
Belgium	6	Mean	24,327	2,012	0.0465%	72.13%
		Min.	843	86	0.0181%	62.44%
		Max.	118,942	8,134	0.0956%	88.78%
France	15	Mean	7,957	1,632	0.0362%	74.73%
		Min.	195	2	0.0063%	62.35%
		Max.	55,979	12,658	0.1006%	97.30%
Germany	13	Mean	10,039	1,997	0.0962%	74.80%
		Min.	242	10	0.0084%	59.25%
		Max.	71,713	15,074	0.4480%	95.63%
Ireland	4	Mean	4,551	291	0.0450%	86.20%
		Min.	1,599	46	0.0010%	79.97%
		Max.	7,898	709	0.0951%	93.07%
Italy	11	Mean	6,495	1,454	0.0305%	86.84%
		Min.	292	7	0.0015%	79.01%
		Max.	27,628	6,234	0.1609%	98.31%
Portugal	4	Mean	6,035	944	0.0047%	74.92%
		Min.	2,080	612	0.0010%	63.14%
		Max.	10,857	1,090	0.0135%	85.44%
Spain	12	Mean	9,650	1,884	0.0098%	85.02%
		Min.	801	299	0.0024%	78.77%
		Max.	40,712	10,613	0.0238%	92.35%
The Netherlands	11	Mean	7,747	1,771	0.0181%	75.43%
		Min.	383	54	0.0014%	64.80%
		Max.	50,233	9,036	0.0607%	87.64%
The United Kingdom	15	Mean	8,529	1,228	0.0189%	65.27%
		Min.	395	16	0.0028%	53.47%
		Max.	69,843	6,969	0.0480%	79.80%
Total	91	Mean	9,481	1,468	0.0340%	77.26%
		Min.	195	2	0.0010%	53.47%
		Max.	118,942	15,074	0.4480%	98.31%

This table reports the number of stocks sampled by country and, for each country, the average, the minimum, and the maximum values of the market value in million euros, the total traded value in May 2013 in million euros, the cross-market bid-ask spread, and the market share of the primary exchange. Four markets are considered: the primary exchange, Chi-X, Bats, and Turquoise.

Table 2. Member categories

Trading scope	Trading aggressiveness	Trading speed	Capacity	Number of member/stock combinations	% in trading volume				
					Total	Primary exchange	BATS	Chi-X	Turquoise
Local trader	Liquidity taker	Slow	A	3,259	15.80%	15.72%	0.01%	0.06%	0.01%
			P	1,241	4.88%	4.31%	0.02%	0.37%	0.18%
		AT	A	247	3.79%	3.78%	0.00%	0.01%	0.00%
			P	105	0.39%	0.30%	0.00%	0.03%	0.06%
		HFT	P	34	0.35%	0.19%	0.00%	0.16%	0.00%
	Liquidity supplier	Slow	P	545	0.99%	0.81%	0.01%	0.10%	0.07%
		AT	P	122	0.50%	0.36%	0.01%	0.12%	0.02%
		HFT	P	61	0.48%	0.29%	0.01%	0.18%	0.01%
Global trader	Liquidity taker	Slow	A	527	3.23%	1.87%	0.24%	0.89%	0.22%
			P	817	20.22%	11.70%	1.13%	5.27%	2.12%
		AT	A	189	3.18%	1.82%	0.18%	0.63%	0.55%
			P	231	7.37%	4.19%	0.42%	1.59%	1.18%
		HFT	P	305	15.31%	8.34%	0.94%	4.11%	1.93%
	Liquidity supplier	Slow	P	441	9.69%	5.73%	0.57%	2.42%	0.98%
		AT	P	218	7.75%	3.13%	0.64%	2.44%	1.55%
		HFT	P	226	6.06%	1.81%	0.76%	2.54%	0.94%
Total				8,568	100%	64.35%	4.92%	20.91%	9.82%

This table displays the relative market size of each member category. Our member classification is established on a stock-by-stock basis and based on three criteria: local vs. global traders, liquidity suppliers vs. liquidity takers, and slow traders vs. ATs/HFTs. Flags for a given member on a given stock can also differ according to the member capacity (agent or principal). As a result, column "Number of member/stock combinations" displays numbers of member×capacity×stock combinations. The right-hand side of the table reports the percentages of each category in total trading volumes with a breakdown by exchanges.

Table 3. Average level of GL as a percentage of quantities available in the book

		10ms	Refill rate in the next 10ms	20ms	50ms	100ms
Panel A - By country						
Belgium		5.81%	0.35%	5.85%	5.93%	6.13%
France		4.95%	0.25%	5.11%	5.31%	5.39%
Germany		3.04%	-0.35%	3.44%	3.68%	3.73%
Ireland		3.26%	-0.74%	2.59%	2.09%	2.32%
Italy		1.73%	-1.34%	2.05%	2.19%	2.32%
The Netherlands		6.18%	-0.39%	6.34%	6.20%	6.36%
Portugal		3.53%	0.12%	3.60%	3.55%	3.51%
Spain		0.23%	0.06%	0.43%	0.61%	0.50%
The United Kingdom		6.83%	-0.74%	6.92%	6.81%	6.91%
All stocks		4.04%	-0.34%	4.20%	4.26%	4.34%
Panel B - By platform						
Primary exchange		3.43%	-0.55%	3.54%	3.48%	3.54%
Chi-X		6.58%	-0.78%	7.06%	7.47%	7.61%
Turquoise		5.98%	0.58%	6.22%	6.54%	6.72%
BATS		6.92%	-0.54%	7.56%	8.19%	8.51%
Panel C - By pair of platforms						
GL venue	*Trade venue*					
Primary exchange	Chi-X	3.74%	-0.48%	3.87%	3.92%	4.02%
	BATS	1.96%	-0.19%	2.00%	1.69%	1.50%
	Turquoise	3.30%	-0.57%	3.38%	3.34%	3.37%
Chi-X	Primary exchange	6.61%	-0.86%	7.11%	7.58%	7.80%
	BATS	5.25%	-1.03%	5.56%	5.48%	4.97%
	Turquoise	6.31%	-0.31%	6.51%	6.63%	6.60%
BATS	Primary exchange	6.19%	-0.68%	6.82%	7.54%	7.93%
	Chi-X	8.50%	-1.41%	9.39%	9.77%	9.72%
	Turquoise	8.55%	-0.86%	8.79%	9.02%	9.21%
Turquoise	Primary exchange	5.86%	0.65%	6.07%	6.45%	6.73%
	Chi-X	5.99%	-0.33%	6.28%	6.34%	6.30%
	BATS	4.94%	-0.89%	5.13%	5.03%	5.17%

This table reports statistics on GL measured as a percentage of quantities available in the order book prior to executions on the trade venue. Means of GL are presented by country (Panel A), platform (Panel B), and pair of platforms (Panel C), for different time windows (10ms, 20ms, 50ms, and 100ms). For GL at the 10ms horizon, the table also reports the average refill rate within the next 10ms. Refill rates are winsorized at the 99% level. GL and refill rates are first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members are constructed, where the weight for a member is equal to that member's average contribution to order book depth over the day. Finally, those daily weighted values are averaged for each stock over the entire month and equally-weighted means across stocks are calculated.

Table 4. Average level of GL as a percentage of trade size

		10ms	20ms	50ms	100ms
Panel A - By country					
Belgium		20.48%	21.01%	21.67%	23.47%
France		16.69%	17.88%	19.40%	19.72%
Germany		7.96%	9.22%	10.12%	10.48%
Ireland		30.69%	30.29%	27.56%	28.98%
Italy		9.49%	10.97%	13.08%	14.07%
The Netherlands		27.10%	28.35%	29.38%	30.15%
Portugal		19.29%	19.25%	18.71%	18.75%
Spain		0.63%	1.02%	1.43%	1.28%
The United Kingdom		42.18%	44.86%	41.96%	43.47%
All stocks		18.89%	20.11%	20.34%	21.07%
Panel B - By platform					
Primary exchange		20.54%	21.83%	21.61%	22.31%
Chi-X		18.26%	19.54%	21.35%	22.14%
Turquoise		18.62%	19.63%	20.34%	21.54%
BATS		16.82%	18.41%	20.74%	21.31%
Panel C - By pair of platforms					
GL venue	*Trade venue*				
Primary exchange	Chi-X	22.87%	24.11%	22.61%	23.49%
	BATS	19.02%	21.32%	21.91%	22.28%
	Turquoise	19.43%	20.16%	21.13%	21.78%
Chi-X	Primary exchange	16.62%	17.81%	19.53%	20.31%
	BATS	30.67%	32.38%	35.42%	36.10%
	Turquoise	25.91%	27.85%	29.54%	31.25%
BATS	Primary exchange	12.84%	14.41%	16.89%	17.30%
	Chi-X	30.29%	33.05%	34.93%	35.60%
	Turquoise	30.81%	31.69%	33.76%	35.90%
Turquoise	Primary exchange	16.60%	17.48%	18.31%	19.55%
	Chi-X	26.62%	28.38%	29.04%	30.48%
	BATS	27.30%	28.58%	27.38%	30.26%

This table reports statistics on GL measured as a percentage of the size of the trade that triggers the GL measurement. Means of GL are presented by country (Panel A), platform (Panel B), and pair of platforms (Panel C), for different time windows (10ms, 20ms, 50ms, and 100ms). GL is first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members are constructed, where the weight for a member is equal to that member's average contribution to order book depth over the day. Finally, those daily weighted GL values are averaged for each stock over the entire month and equally-weighted means across stocks are calculated.

Table 5. Average level of GL per market value tercile, volatility tercile, and fragmentation tercile

		Average GL as a % of pre-trade liquidity (10ms)	Refill rate in the next 10ms	Average GL as a % of trade size (10ms)
Market value tercile	**Market value range (EUR Mn)**			
1	195 to 1,833	3.45%	0.39%	16.17%
2	1,989 to 5,846	3.86%	-0.50%	17.98%
3	6,152 to 118,942	4.79%	-0.88%	22.42%
Volatility tercile	**Daily volatility range**			
1	0.0706% to 0.1253%	4.96%	-0.52%	22.74%
2	0.1266% to 0.1549%	3.97%	-0.22%	18.78%
3	0.1549% to 0.3266%	3.17%	-0.26%	15.04%
Fragmentation tercile	**Fragmentation index range**			
1	1.0604 to 1.5520	1.68%	0.42%	7.18%
2	1.5553 to 2.0663	3.35%	-0.27%	15.10%
3	2.0831 to 3.0714	7.00%	-1.13%	33.90%

This table reports statistics on GL by market value tercile, volatility tercile, and fragmentation tercile. GL is here measured both as a percentage of the pre-trade quantity posted by a member in the order book as well as in percentage of the size of the trade that triggers the measurement of GL. The table also reports average refill rates within the next 10ms. Refill rates are winsorized at the 99% level. GL and refill rates are first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members are constructed, where the weight for a member is equal to that member's average contribution to order book depth over the day. Finally, those daily weighted values are averaged for each stock over the entire month and equally-weighted means across stocks are calculated.

Table 6. Average GL as a percentage of pre-trade liquidity by member category

		Average GL as a % of pre-trade liquidity (10ms)	% of cases with duplication	Refill rate in the next 10ms	Average GL as a % of trade size (10ms)
Trading aggressiveness	Liquidity Taker	3.69%	34.42%	1.34%	13.53%
	Liquidity Supplier	3.81%	54.84%	-0.02%	18.43%
Trading scope	Local	2.11%	3.31%	0.26%	11.59%
	Global	3.80%	57.81%	0.38%	16.50%
Trading speed	Slow	2.70%	32.60%	0.08%	12.32%
	AT	3.76%	56.84%	0.81%	12.52%
	HFT	5.75%	53.65%	0.09%	16.87%
Capacity	Agent	1.94%	16.78%	3.16%	5.48%
	Principal	3.93%	51.23%	0.42%	17.56%

This table reports statistics on GL and refill rates by member category. It also provides the proportion of trades for which pre-trade liquidity is duplicated. Refill rates are winsorized at the 99% level. GL and refill rates are first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members in the considered category are constructed, where the weight for a member is equal to that member's average contribution to order book depth over the day. Finally, those daily weighted values are averaged for each stock over the entire month and equally-weighted means across stocks are calculated.

Table 7. Tobit regressions of GL for global market members

GL measure	GL as fraction of pre-trade liquidity				GL as fraction of trade size
Time window for GL	10ms	20ms	50ms	100ms	10ms
Member characteristics					
HFT	0.0788***	0.0788***	0.0807***	0.0826***	0.2197***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
AT	0.0280***	0.0279***	0.0284***	0.0287***	0.0547***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Agent	-0.0203***	-0.0198***	-0.0210***	-0.0231***	-0.0522***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Liquidity supplier	0.0258***	0.0275***	0.0280***	0.0281***	0.0851***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average inventory $_{t-1}$	-0.0009***	-0.0008***	-0.0009***	-0.0010***	-0.0041***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Trade characteristics					
Trade size $_{t}$	0.0119***	0.0119***	0.0120***	0.0119***	
	(0.000)	(0.000)	(0.000)	(0.000)	
Platform characteristics					
PE-to-alternative	-0.0183***	-0.0170***	-0.0149***	-0.0147***	-0.0611***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Alternative-to-PE	0.0267***	0.0269***	0.0275***	0.0284***	0.1100***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Other market member GL					
GL(other, HFT) $_{t}$	0.0472***	0.0480***	0.0497***	0.0495***	5.28E-06***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
GL(other, AT) $_{t}$	0.0540***	0.0458***	0.0513***	0.0392***	3.48E-05***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GL(other, Slow) $_{t}$	0.0000	0.0000	0.0000*	0.0000*	1.39E-05***
	(0.167)	(0.118)	(0.099)	(0.083)	(0.000)
Order flow characteristics					
Volume $_{t}$	0.0047***	0.0050***	0.0051***	0.0051***	0.0165***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Order imbalance $_{t}$	-0.0092***	-0.0098***	-0.0098***	-0.0106***	-0.0383***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Order imbalance $_{t-1}$	-0.0009	-0.0009	-0.0004	-0.0005	0.0009
	(0.161)	(0.158)	(0.570)	(0.436)	(0.626)
Fragmentation $_{t-1}$	0.0020***	0.0023***	0.0024***	0.0023***	0.0060***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

SOR $_{t-1}$	0.3755***	0.3873***	0.3989***	0.4122***	1.2192***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$(SOR_{t-1})^2$	-0.8473***	-0.8871***	-0.9174***	-0.9625***	-3.1961***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Stock characteristics					
Price range $_{t-1}$	0.0266***	0.0233***	0.0223**	0.0238**	0.0473*
	(0.007)	(0.024)	(0.039)	(0.032)	(0.097)
Price	-0.0053	-0.0022	-0.0051	-0.0027	0.0092
	(0.179)	(0.592)	(0.223)	(0.531)	(0.443)
Tick	-15.1442***	-14.2076***	-12.3683**	-8.1827	-128.6595***
	(0.006)	(0.012)	(0.034)	(0.169)	(0.000)
Fixed effects					
stock fixed effects	YES	YES	YES	YES	YES
15-min period fixed effects	YES	YES	YES	YES	YES
Pseudo R^2	9.69%	9.06%	8.72%	8.44%	8.35%

This table reports the conditional marginal effects estimated from Tobit regressions of 15 minute GL by member, stock, and pairs of platforms on various dummy variables and other controls. GL is computed in several ways, first as a fraction of pre-trade liquidity on the quote venue over four different time intervals (10ms, 20ms, 50ms, and 100ms) and then as a fraction of trade size at the 10ms horizon. GL is computed only using trades of global members. Each pair of platforms consists of the trade venue, i.e., the venue where the member was passively executed, and the GL venue, i.e., the venue where the member's liquidity is potentially withdrawn. Reported coefficients are the marginal effects of the explanatory variables on GL, conditional on GL being positive. The control variables include a measure of daily realized volatility; the imbalance between buy and sell orders as a percentage of the total traded volume; the log of the total daily traded volume; the log of the closing price; the relative tick size; the contemporaneous GL measured for other HFT members; contemporaneous GL measured for other AT members; contemporaneous GL measured for other slow traders; a fragmentation index; the average size of the trades triggering the GL observation; an HFT dummy equal to one for HFT members; an AT dummy equal to one for AT members; an agent dummy equal to one for a member trading as agent; a liquidity-supplier dummy equal to one for members identified as liquidity providers; a PE-to-ALT dummy equal to one when the trade venue is the primary exchange and the GL venue an alternative platform; a ALT-to-PE dummy equal to one when the trade venue is an alternative platform and the GL venue is the primary exchange. When GL is measured as a fraction of pre-trade quantities in the book of the quote venue the Tobit specifications are double-censored with a lower bound set to 0 and an upper bound set to 1. GL as a percentage of trade size is winsorized at the 99% level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level respectively.

Table 8. Tobit regressions of GL as a percentage of pre-trade liquidity by member sub-samples

	(1)	(2)	(3)	(4)	(5)
	Fast traders only	**ATs only**	**HFTs only**	**Liquidity suppliers only**	**Fast liquidity suppliers only**
Member characteristics					
HFT	0.0551***			0.0782***	0.0465***
	(0.000)			(0.000)	(0.000)
AT				0.0332***	
				(0.000)	
Agent	-0.0161***	-0.0075***			
	(0.000)	(0.000)			
Liquidity supplier	0.0280***	0.0440***	0.0199***		
	(0.000)	(0.000)	(0.000)		
Average inventory $_{t-1}$	-0.0020***	-0.0016***	-0.0026***	-0.0011***	-0.0023***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Trade characteristics					
Trade size $_{t}$	0.0112***	0.0157***	0.0045***	0.0110***	0.0106***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Platform characteristics					
PE-to-alternative	-0.0240***	-0.0132***	-0.0363***	-0.0080***	-0.0060***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Alternative-to-PE	-0.0033***	0.0284***	-0.0430***	0.0137***	-0.0178***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Other market member GL					
GL(other, HFT) $_{t}$	0.0810***	0.0022***	0.1953***	0.0495***	0.0775***
	(0.000)	(0.008)	(0.000)	(0.000)	(0.000)
GL(other, AT) $_{t}$	0.0870***	0.1530***	0.0060***	0.0498***	0.0757***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GL(other, Slow) $_{t}$	0.0000	0.0155***	0.0000	0.0000	0.0000
	(0.632)	(0.000)	(0.134)	(0.852)	(0.114)
Order flow characteristics					
Volume $_{t}$	0.0058***	0.0075***	0.0047***	0.0030***	0.0033***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Imbalance $_{t}$	-0.0084***	-0.0060***	-0.0113***	-0.0089***	-0.0069***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Imbalance $_{t-1}$	-0.0036***	-0.0040***	-0.0044***	-0.0013*	-0.0022**
	(0.000)	(0.001)	(0.001)	(0.089)	(0.026)
Fragmentation $_{t-1}$	0.0035***	0.0046***	0.0034***	0.0013***	0.0024***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

SOR $_{t-1}$	0.3389***	0.4541***	0.1731***	0.3948***	0.3193***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$(SOR_{t-1})^2$	-0.9354***	-1.1113***	-0.5573***	-0.9429***	-0.8731***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Stock characteristics					
Price range $_{t-1}$	0.0377***	0.0490**	0.0356**	0.0300**	0.0240
	(0.005)	(0.014)	(0.042)	(0.020)	(0.121)
Price	-0.0033	-0.0239***	0.0191**	0.0094*	0.0063
	(0.532)	(0.001)	(0.014)	(0.062)	(0.309)
Tick	3.0240	-5.3349	5.8341	-31.6777***	-18.8915**
	(0.640)	(0.589)	(0.507)	(0.000)	(0.017)
Fixed effects					
stock fixed effects	YES	YES	YES	YES	YES
15-min period fixed effects	YES	YES	YES	YES	YES
Pseudo R^2	9.88%	9.45%	12.84%	10.86%	12.15%

This table reports the conditional marginal effects estimated from Tobit regressions of 15 minute GL by member, stock, and pairs of platforms on various dummy variables and other controls. GL is computed as a fraction of pre-trade liquidity on the quote venue over 10ms windows. GL is computed only using trades of global members. The Tobit regressions are run for five different subsamples of members with double-censoring, the lower bound being set to 0 and the upper bound being set to 1. Each pair of platforms consists of the trade venue, i.e., the venue where the member was passively executed, and the GL venue, i.e., the venue where the member's liquidity is potentially withdrawn. Reported coefficients are the marginal effects of the explanatory variables on GL, conditional on GL being positive. The control variables include a measure of daily realized volatility; the imbalance between buy and sell orders as a percentage of the total traded volume; the log of the total daily traded volume; the log of the closing price; the relative tick size; the contemporaneous GL measured for other HFT members; contemporaneous GL measured for other AT members; contemporaneous GL measured for other slow traders; a fragmentation index; the average size of the trades triggering the GL observation; an HFT dummy equal to one for HFT members; an AT dummy equal to one for AT members; an agent dummy equal to one for a member trading as agent; a liquidity-supplier dummy equal to one for members identified as liquidity providers; a PE-to-ALT dummy equal to one when the trade venue is the primary exchange and the GL venue an alternative platform; a ALT-to-PE dummy equal to one when the trade venue is an alternative platform and the GL venue is the primary exchange. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level respectively.

Table 9. Daily effective spreads and GL

	Effective spreads of slow LTs	Effective spreads of ATs/LTs	Effective spreads of HFTs/LTs	Effective spreads of slow LTs	Effective spreads of ATs/LTs	Effective spreads of HFTs/LTs
Realized volatility	0.0717***	0.0557***	0.0415***	0.0711***	0.0528***	0.0414***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Daily volume	-3.61E-05***	-2.75E-05***	-2.73E-05***	-4.18E-05***	-2.33E-05***	-2.77E-05***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Closing price	-0.0001**	-0.0002***	-0.0002***	-0.0001	-0.0002***	-0.0002***
	(0.031)	(0.002)	(0.007)	(0.188)	(0.005)	(0.007)
Average trade size	8.75E-06***	8.14E-06***	1.18E-06	6.07E-06***	7.55E-06***	1.03E-06
	(0.000)	(0.002)	(0.634)	(0.001)	(0.003)	(0.677)
Primary exchange	6.61E-05***	2.27E-05***	2.03E-05***	5.21E-05***	2.11E-05***	2.16E-05***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GL	1.22E-08	1.33E-07***	1.01E-07			
	(0.697)	(0.000)	(0.174)			
GL×primary exchange	1.99E-06***	9.28E-08	1.07E-06			
	(0.000)	(0.879)	(0.136)			
GL of HFTs				-1.10E-05***	-1.72E-06	1.56E-06
				(0.000)	(0.496)	(0.767)
(GL of HFTs)×primary exchange				5.38E-05***	-9.27E-06	7.85E-06
				(0.000)	(0.171)	(0.369)
Average effective spread on the previous day	0.1879***	0.1729***	0.2011***	0.2013***	0.2296***	0.2010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fixed effects						
Stock fixed effects	YES	YES	YES	YES	YES	YES
Day fixed effects	YES	YES	YES	YES	YES	YES
Adjusted R^2	71.23%	69.71%	67.43%	74.16%	70.40%	67.42%

This table reports the OLS regression results of daily effective spreads on GL measures. The three columns consider the daily effective spreads for each stock-venue combination for slow liquidity takers, algo liquidity takers, and HFT liquidity takers as dependent variable, respectively. GL is the ghost liquidity for that stock on that venue, GL of HFTs is the ghost liquidity for that stock on that venue stemming from HFTs, and primary exchange is a dummy indicating whether the observation stems from the primary exchange or not. The control variables include a measure of daily realized volatility, the daily volume, the closing price, average trade size, and the lagged effective spread for that stock and trader group. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level respectively.