Duplicated Orders, Swift Cancellations, and Fast Market Making in Fragmented Markets*

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

May 15, 2023

JEL classification: G14, G15, G18

Keywords: High Frequency Trading (HFT), Algorithmic Trading (AT), Fragmentation, Limit order duplication, Order cancellation

* We would like to thank Panos Anagnostidis, Carlos Aparicio Roqueiro, Jos van Bommel, Antoine Bouveret, Carole Comerton-Forde, Sarah Draus, Thierry Foucault, 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, members of the AMF scientific board, participants at the 35th Spring AFFI conference, the 4th WiM meeting, the 2nd ECM Workshop, the 2nd SAFE Market Microstructure Conference, the 2018 FMA annual meeting, the 46th EFA meeting (Lisbon), and the 2020 Plato Market Innovator (MI3) conference for their comments. We thank Yujuan Zhang for excellent research assistance, ESMA for providing access to the data, the French National Research Agency (ANR) for funding the project, and CEPREMAP for granting additional research budget. The views presented in this paper are those of the authors and do not necessarily reflect the views of ESMA.

Duplicated Orders, Swift Cancellations, and Fast Market Making in Fragmented Markets

Abstract

Employing unique data from 91 stocks trading on their primary exchanges and three alternative venues, we show that limit orders are duplicated across competing venues as a result of cross-venue market-making strategies, leading to swift cancellation of duplicate orders after one of them is filled. This Duplicated-then-Canceled Liquidity is predominantly used by high-frequency traders when their inventories are not excessive. It reduces the execution costs of fast traders on alternative venues. It however has some adverse impact on execution costs on primary exchanges but those negative effects fail to outweigh the liquidity benefits of market fragmentation.

1. Introduction

Recent developments in equity market structure have rendered the process of identifying and implementing optimal execution and market-making strategies significantly more complex. First, the fragmentation of modern equity markets and the use of multiple trading venues by market participants means that, to understand liquidity, one must aggregate across many venues and data feeds to obtain a 'consolidated' view of the market, while to execute efficiently often requires the use of a 'smart order router' (see, for example, Foucault and Menkveld, 2008). Second, the same market developments have led to changes in traders' limit 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 the average impatient market participant can access. This is because, in a world of fragmented market order flow, traders implementing market-making strategies may rationally choose to place duplicate limit orders on several venues, intending for only one of those orders to execute. In this work, we empirically evaluate the extent of this order duplication, how it is related to cross-venue market-making activity, and we quantify the extent to which consolidated liquidity overstates true liquidity.

To illustrate the key issue, 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 creates a difference between true and measured liquidity. 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 duplicate limit buy order submitted by the patient trader 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 canceled. As a result, the liquidity actually accessible to her is less than initially observed. We call this difference *Duplicated-then-Canceled Liquidity* (DCL).

To be clear, we are not defining DCL to arise from orders which were never intended to execute under any circumstance, but to arise due to orders which are canceled conditional on the execution of a similar order submitted by the same trader on another venue. This means that DCL is not necessarily inaccessible to all traders. One may fairly claim that DCL is genuine liquidity between its submission time and its cancellation time, not only for fast multi-market aggressive traders, but also for the small local traders of all venues. As the traders who operate cross-venue market-making strategies must take positions in order to generate profits, DCL cannot be entirely irrelevant, and if it contributes to more effective inventory management for these traders it is likely to make the consolidated market more liquid. Notwithstanding this, a significant fraction of DCL may turn out to be unavailable to the average trader most of the time.

It is worth noting that order duplication is not without risk. If both of the passive trader's limit buy orders in our example 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 and simultaneously executes the passive trader's orders on both venues. 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 duplication of orders across venues by fast traders may be a natural part of cross-venue market-making, but it may also lead measured liquidity to overstate true liquidity for the average trader. The core of this paper is an attempt to quantify the importance of DCL in equity markets and to characterize its determinants. We take advantage of a unique dataset 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 DCL 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 DCL as a fraction of the size of the triggering execution and as a fraction of total liquidity. We then regress DCL measures on a set of trader characteristics, venue characteristics, and exogenous variables to characterize the determinants of duplication. Last, we analyze the impact of DCL on the trading costs of slow and fast liquidity takers.

We find that DCL accounts for a sizeable fraction of order cancellation activity. To a first 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. However, with an average level of around 4% of the total cross-market consolidated depth, it is unlikely that DCL is large enough to outweigh the liquidity benefits of market fragmentation. DCL is still worthy of measurement as its cross-sectional and time-series variation lead to it being substantial in some situations and because understanding its determinants sheds light on how liquidity suppliers conduct their market-making and inventory management strategies in fragmented markets. There are variations across venues and countries, with DCL 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 canceled on a competing venue on average). Duplication is more prevalent 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 duplication also shows that trader characteristics are important. High-frequency traders (HFTs) have the largest measures of DCL, followed by algorithmic trading (AT) firms. Traders who are neither HFTs nor ATs, a group that we call 'slow' traders, have the lowest DCL levels. Duplication is also more common when a trader is acting as a principal rather than as an agent. This supports the idea that DCL is an important part of high-speed market-making strategies.

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 duplicated orders most heavily when other traders are also doing so and that DCL increases when stock-specific trading volumes are high. We find that when the execution that triggered the duplicate order removal was large, the fraction of displayed liquidity that a trader removes increases. There is also evidence that DCL effects are strongest when the triggering trade is on an alternative trading venue (i.e., not the primary exchange of the traded stock) and when the venue where liquidity is being removed is also an alternative trading venue. 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 the posting, and subsequent cancellation, of duplicated orders. Finally, we find that when the inventory level of the liquidity supplier is large, she is marginally less likely to duplicate orders. This leads us to reject the hypothesis that duplication is used as a tool for rebalancing extreme inventories. Instead, it appears that traders are more comfortable to submit orders to multiple venues when inventories are not excessive. More explicitly, this indicates that limit order traders implementing liquidity-supplying strategies are more prone to use DCL in the first phase of those strategies, when they build trading positions, than in the second phase, when they unwind them. We also find that when the prevalence of smart order routing is particularly large, it tends to reduce DCL. This result is reasonable 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 DCL to a study of its implications, focusing on the DCL activity generated by HFTs. We examine whether the level of duplication impacts upon execution costs, measured by effective spreads, that various trader groups pay. We find that DCL on primary venues leads to increases in slow trader and algo trader (AT) execution costs. The effect is economically and statistically small for slow traders, though. On the other hand, there is strong evidence that the DCL activity of HFTs reduces the trading costs of algo traders on alternative trading venues. Finally, DCL reduces HFT trading costs on all venues. We interpret this as implying that DCL forms an important part of HFT market-making strategies which allows their counter-parties access to greater liquidity on non-primary trading venues.

Overall, our results show that the use of order duplication is part of market-making strategies employed by fast trading firms in fragmented markets and, in particular, when their inventory levels are not excessive. Therefore, it contributes to the formation of liquidity on all venues, with greater benefits on alternative ones. It however raises questions about the use of simple consolidated liquidity measures to assess market quality. Regarding this matter and the potential adverse effects of DCL, we find that higher DCL is associated with greater execution costs for less sophisticated traders. The economic significance of this effect as well as the cross-market level of DCL are not, however, very large and are unlikely to be sufficiently great to challenge the benefits of fragmentation.

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 measurements of DCL. Section 6 contains our analysis of the determinants of DCL. We examine the impact of DCL 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 Duplicated-then-Canceled Liquidity (DCL). By DCL, we mean liquidity that is supplied to markets but which is 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 several venues. What are the incentives to post multiple orders? In fragmented markets, 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 DCL 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 (2021), 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 to 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.

DCL 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. We will add to this literature by showing that, in fragmented markets, individual traders place duplicate orders on several exchanges. This challenges the result that fragmentation leads to larger *measured, consolidated* liquidity as, with duplicated limit orders, *measured* and *real* liquidity may 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 canceled 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, a reasonable market-making strategy may involve the posting of duplicate limit orders on more than one venue, only intending for one of the orders to execute and (partially) canceling the duplicates once an execution occurs. The latency advantage enjoyed by these traders who run market-making strategies 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 DCL and which implies that measured, consolidated liquidity is larger than real liquidity. It is worth noting, though, that, to the extent that it makes inventory management easier for those operating market-making strategies, it is likely to improve real liquidity.

Our definition of DCL above suggests that looking at order duplication and order cancellations on one venue in response to trades on another might be useful in identifying DCL. 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.[1] 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.

However, van Kervel (2015) shows that in worlds with no DCL (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 HFTs acting as market-makers post quotes on all venues simultaneously. In the absence of any new information those HFTs 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 reflect real depth and not duplicated

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

orders. 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 resulting from order duplication but it is rational updating of quotes in response to new information. Thus, one observes cross-venue cancels in a world without DCL. 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 DCL 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 DCL and its role in cross-venue market making 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.[2] We are therefore able to make four important contributions to the literature. First, we estimate the importance of DCL for a given trader. Second, we compare the importance of DCL across different groups of traders, and across different venues. Third, based on our measurement of DCL by trader, we identify its economic determinants and can, through this, shed light on how cross-venue market-making strategies operate. Last, we assess the impact of DCL 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 historically main exchanges of nine countries comprising Belgium, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, and the United Kingdom. Those national exchanges will be referred to as "primary" exchanges and denoted *PE* in the empirical analysis.[3] The dataset covers the trading and quoting activity on the primary exchanges and on the three largest alternative exchanges on which the sample stocks were admitted for trading at that time, namely

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

BATS, Chi-X and Turquoise.[4] These three venues will be referred to as "alternative" exchanges and denoted *ALT* in the rest of the paper. Together with the primary exchanges, they deal with the vast majority of trading activity for each stock. Bouveret et al. (2014) were the first to employ the dataset in their analysis of the extent of HFT in European stock markets. Further details on the construction and content of the dataset can be found there.

All exchanges in our study are regulated under the Markets in Financial Instruments Directive (MiFID). Key provisions in MiFID include the abolition of trading concentration on main exchanges, pre-trade and post-trade transparency requirements for all trading venues, and best execution rules. A key difference between MiFID and Reg NMS in the U.S. is that MiFID defines best execution as not only a matter of the price of a trade, but also "costs, speed, likelihood of execution and settlement, size, nature or any other consideration relevant to the execution of the order" (Art. 21). While Rule 602b of Reg NMS obliges trading venues to execute orders at the best price or to re-route them to the venue quoting the best price in order to prevent trade-throughs, MiFID has softer requirements. As a consequence, European exchanges do not route orders to each other, and order flow fragmentation might distort both time priority and, occasionally, price priority.

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. To account for the relative size of the markets, a 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

[4] At the current date, Chi-X and BATS, renamed CXE and BXE respectively, are operated by company CBOE Global Markets as a result of CBOE taking over Bats Global Markets in 2017.

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 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) to 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.

[5] Either those stocks were not traded over several days or they were not traded outside the primary exchange.

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, definitions are still evolving and the academic literature adopts a variety of methods 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 either HFT or non-HFT based on public information 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 a set of fast traders using the indirect approach of Bouveret et al. (2014) based on the lifetime of orders.[6]

Bouveret et al. (2014) use the same data as we do and they classify members as fast traders if the 10% quickest order modifications and cancellations in a given stock occur no more than 100ms

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

after the initial submission.[7] Such a criterion indicates that the member under consideration possesses fast trading technology even if she does not use it at all times. We follow Bouveret et al. (2014) in choosing 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. It is worth noting that Bouveret et al. (2014) find that just over 40% of value traded is done by fast traders using this approach.[8]

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.

Second, we subdivide the population of fast traders into two distinct groups, HFTs and other algorithmic traders. This second identification, based on the direct approach of Bouveret et al. (2014), results in a list of 21 HFT firms. This list is built using firms' websites and the financial press to identify 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 then define algorithmic traders (ATs) as the residual subset of fast traders who are not 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 while 81 trade on multiple platforms. The distinction between

[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.

[8] 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.

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 DCL is defined as a consequence of multi-market trading strategies. We therefore classify global members as market participants who trade on 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

Duplicated limit orders are 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, DCL can only be generated by traders implementing market making strategies. For that reason, it seems relevant 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 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 who is 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 dataset via the previously mentioned capacity flag, resulting in distinct member/capacity pairings for that member and that stock. While Bouveret et al. (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. 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.

ATs and HFT firms account for respectively 22.98% and 22.21% of the total traded value. 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. Lastly, liquidity suppliers account for 25.47% of the total traded value.

4.5. Market fragmentation level and investor clientele segmentation

In terms of trading volume distribution, as reported in Table 2, 64.35% of the total volume is traded on primary exchanges, and Chi-X is the main alternative venue with a volume share of 20.91%. All platforms except BATS have venue-specific local traders. Of our 388 market members, 307 trade on only one venue, with 297 trading only on the primary exchange, eight trading only on Chi-X and two only on Turquoise. Those 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. Those figures indicate that the market architecture is substantially fragmented, with more than a third of lit trading volumes executed outside primary exchanges, and with single-venue trading occurring on every venue. However, this level of fragmentation is less than that observed in U.S. stock markets, where the share of primary exchanges is less than 40%. This fragmentation level, which is intermediate between consolidation and total fragmentation, is referred to as market segmentation in Harris (1993).

In segmented markets, global traders in general, and global liquidity suppliers in particular, who respectively represent 72.81% and 23.50% of the total traded volume in our sample, play a crucial role in partially reconsolidating the order flow and synchronizing prices. Among our 388 market members, 81 members trade on multiple platforms: 39 trade on all four platforms, 17 trade on three platforms only, and 25 trade on two platforms only. 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 statistics of Table 2 also show that primary exchanges and alternative venues attract different trader clienteles. First, fast traders are much more present on alternative venues. The relative weight of ATs and HFTs is greater on BATS, Chi-X, and Turquoise, where their respective volume shares are 26.40% and 32.47%, versus 21.09% and 16.53% respectively on primary exchanges. BATS and Chi-X have the highest shares of volume traded by HFTs (respectively 34.83% and 33.40% of their traded volumes) while Turquoise has the highest share of volume traded by ATs (34.14%). Second, the weight of global traders is greater on alternative platforms, where they account for 96.02% of the volumes. On the contrary, primary exchanges have the highest level of local trading volumes (25.77% versus 4.90% on Chi-X and 3.52% on Turquoise). Third, liquidity suppliers are more active on alternative venues, where they trade 37.45% of the volume.

In markets segmented both in terms of order flow and trader type, understanding to what extent and how duplicated limit orders are used in liquidity providing strategies is of great interest. On the one hand, the swift cancellation of duplicates may adversely impact the execution costs of some traders, but on the other hand, limit orders duplicated across venues may be a source of liquidity for the local traders of all venues.

5. Assessing the level of Duplicated-then-Canceled Liquidity (DCL)

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 DCL at different levels of aggregation (trader, venue, …). Subsection 5.1 describes the methodology we use to measure DCL and to aggregate it at different levels. Subsection 5.2 describes how we check whether the DCL we measure is genuine reduction in depth or whether it is immediately followed by re-supply of liquidity by the same trader on the same side of the market but at a different price point, thereby suggesting quote updating. In Subsection 5.3, we then investigate whether DCL is immediately followed by liquidity provision by the same trader on the opposite side of the market, reflecting market-making activity. Subsection 5.4 reports descriptive statistics.

5.1. Measuring DCL

Our DCL metric is based on the following simple intuition. Assume that a trader is posting limit sell orders, for example, on several venues simultaneously. Also assume 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 canceled then those canceled orders represented DCL.

As the simple example above makes clear, DCL 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 canceled in response to an execution of another of that trader's orders on a different venue to qualify as DCL? How similar does the canceled order have to be to the executed order to count as DCL? Any definition of DCL will have to be flexible enough to take account of all of the above.

We begin with a specification of DCL 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 DCL posted by m on venue qv is equal to:

$$DCL_{tv \to qv}^{ask}\left(\tau;\Delta\tau;i;m\right) = PREQTY_{qv}^{ask}\left(\tau;i;m\right) - POSTQTY_{qv}^{ask}\left(\tau;\Delta\tau;i;m\right) - \sum_{\tau;\Delta\tau} Volume_{qv}^{buy}\left(i;m\right) \qquad (1)$$

where $PREQTY_{qv}^{ask}\left(\tau;i;m\right)$ 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 $POSTQTY_{qv}^{ask}\left(\tau;\Delta\tau;i;m\right)$ 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 DCL.

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.[9] 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 DCL measure. Assume that we are measuring DCL 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

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

current midquote such that only orders within that band contribute to the DCL measure. We have chosen the width of the band 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 DCL measurement.

The baseline DCL measure above is trader, trade time, stock, venue, and side specific, and we want to aggregate these data so that they can be compared across stocks and times. To make the data comparable across stocks, and to aggregate up to the daily level, we express DCL as a proportion of displayed quantity. First, we compute the following measure:

$$DCL_{tv\to qv}\left(\Delta\tau;i;d;m\right)=\frac{\sum_{\tau\in d} DCL^{bid}_{tv\to qv}\left(\tau;\Delta\tau;i;m\right)+\sum_{\tau\in d} DCL^{ask}_{tv\to qv}\left(\tau;\Delta\tau;i;m\right)}{\sum_{\tau\in d} PREQTY^{bid}_{qv}\left(\tau;i;m\right)+\sum_{\tau\in d} PREQTY^{ask}_{qv}\left(\tau;i;m\right)}. \qquad (2)$$

In Equation (2), trade-time measures $DCL_{tv\to qv}\left(.\right)$ and $PREQTY_{qv}\left(.\right)$ are summed for all trades within a given day to give aggregated DCL 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 DCL 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 DCL are computed for each stock.

Equation (2) gives the DCL 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 because of a trade on another venue. An alternative way to scale DCL 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 DCL 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 DCL measurement. In our empirical work, we perform all estimations using both DCL measures that scale by depth and using DCL measures that scale by trade size.

In our summary statistics we present cross-stock averages of DCL per pair of venues. For each pair of venues, the average computed reflects the mean level of DCL 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 DCL 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 10ms immediately after duplicated order cancellations

One may argue that our DCL measure is not necessarily capturing the cancellation of duplicated limit orders 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 quote venue's order book. No such resubmissions should occur in the case of genuine DCL. Thus, to distinguish DCL from quote updating, we compute a book refill rate for the 10ms after the time window over which DCL is measured. For a given member whose order cancellation has contributed to our DCL calculation, this refill rate equals the liquidity added by that same member on the same venue where DCL is being measured.[10] To be more explicit about the calculation of the refill rate, let us return to the example we used when discussing the DCL 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 DCL 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:

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

$$\left.\left(\begin{aligned}&POSTQTY^{ask}_{qv}\left(\tau+10ms;i;m\right)-PREQTY^{ask}_{qv}\left(\tau+10ms;i;m\right)\\&\qquad\qquad\qquad\qquad+\sum_{\tau+10ms}Volume^{buy}_{qv}\left(i;m\right)\end{aligned}\right)\middle/ DCL^{ask}_{tv\rightarrow qv}\left(\tau;10ms;i;m\right)\right. \quad (3)$$

[10] 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 DCL computation and the midquote we use is that observed at the end of the DCL measurement window.

where $PREQTY_{qv}^{ask}\left(\tau+10ms;i;m\right)$ 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 $POSTQTY_{qv}^{ask}\left(\tau+10ms;i;m\right)$ is the total limit sell order quantity posted by member m on that same venue 10ms later. $\sum_{\tau+10ms} Volume_{qv}^{buy}\left(i;m\right)$ 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 venue book immediately after the duplicated order cancellations. This is then expressed as a percentage of the 10ms DCL 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 canceling orders whereas a negative refill rate indicates that the members continued canceling liquidity after the end of the DCL window. Those refill rates are computed for all trades which generated positive DCL and are then averaged across time, members, and stocks, by countries, platforms, stock terciles, and member categories.

5.3. Measuring opposite-side liquidity provision in the 10ms immediately after duplicated order cancellations

In the same spirit, we look at whether DCL is followed by liquidity provision by the same member m over the next 10ms, on the same venue qv but on the opposite side of the market. We do so to check whether DCL could signal the turning point of a market-making strategy. In other words, the cancellations would happen when the trader stops providing liquidity on one side and switches to the other side of the market. Returning to the example used to illustrate Equations (1) and (3), after measuring sell-side DCL by member m on venue qv over a given 10ms time window, we build a measure of liquidity provision by the same member m on the buy side of venue qv over the next 10ms window as follows:

$LiqProv_{tv\rightarrow qv}^{bid}\left(\tau+10ms;i;m\right)$

$$\left(\begin{array}{r} POSTQTY_{qv}^{bid}\left(\tau+10ms;i;m\right)-PREQTY_{qv}^{bid}\left(\tau+10ms;i;m\right) \\ +\sum_{\tau+10ms} Volume_{qv}^{sell}\left(i;m\right) \end{array}\right) \Big/ DCL_{tv\to qv}^{ask}\left(\tau;10ms;i;m\right) \quad (4)$$

where $LiqProv_{tv\to qv}^{bid}\left(\tau+10ms;i;m\right)$ is the depth added by member *m* on the bid-side of *qv* in the 10ms following her cancellation of duplicated limit orders on the ask side for a quantity of $DCL_{tv\to qv}^{ask}\left(\tau;10ms;i;m\right)$. The measure is then expressed as a percentage of $DCL_{tv\to qv}^{ask}\left(\tau;10ms;i;m\right)$. We refer to this measure as the opposite-side liquidity provision rate. Opposite-side liquidity provision rates are computed for all trades which generated positive DCL and are then averaged across time, members and stocks, by countries, platforms, and pairs of platforms.

5.3. Descriptive statistics for DCL

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

Table 3 about here

Table 3 exhibits DCL statistics as a percentage of pre-trade liquidity. Panel A reports DCL by country and is obtained by averaging across primary exchange and alternative venues. This panel reveals some country-level heterogeneity with DCL varying between 0.23% and almost 7%. The countries with the highest DCL are the Netherlands, the UK, and Belgium whereas Spain and Italy exhibit much lower DCL. These countries with low levels of DCL correspond to countries where order flow fragmentation, and more specifically trading outside the primary exchange, are very limited. Panel A also indicates that the average level of DCL for each country does not change much as one moves from a 10ms DCL measurement window to a 100ms window.

Further, Panel A shows that the average refill rate for all stocks combined is negative. Average refill rates by countries are also either negative or close to zero. This suggests that our DCL measure is not contaminated by cancellations due to members repricing orders in response to trades on other venues.

Turning to the opposite-side liquidity provision rates, we find that they exhibit a very different pattern. Although relatively low in value, with averages by country ranging from 0.20% to 1.17%, they are always strictly positive. We interpret this as indicating that DCL, in part, is related to the practice of cross-venue market making strategies.

Panel B reports DCL by platform, by taking, for a given quote venue (where DCL is measured) the average DCL weighted across the trades that trigger our measurements. We find that DCL is much smaller on primary exchanges than on the three alternative venues.

Panel C breaks down DCL by pairs of venues. The first column of the table gives the name of the venue where DCL is being measured (the quote venue) and the second column gives the name of the venue where the trade that triggers the measurement occurred (the trade venue). 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 highest levels of DCL are observed for pairs of alternative venues where the quote venue is BATS and the trade venue is either Chi-X or Turquoise. As in Panel A, the average value of the refill rate is negative for almost all pairs of platforms, while the average opposite-side liquidity provision rates by pairs of platforms are all positive. Pairs of alternative venues, more specifically those involving Chi-X, exhibit the highest opposite-side liquidity provision rates. This again suggests that DCL plays a role in cross-venue market making strategies, especially on alternative venues.

Table 4 about here

Table 4 has the same structure as Table 3, except it reports figures based on DCL as a fraction of trade size rather than quantity outstanding. These figures measure how many shares a passive trader cancels on one venue, on average, after being executed for 100 shares on another venue. Quantifying DCL this way leads to its mean rising 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 canceled on a different venue by the same trader. The UK and the Netherlands are still among the countries with the highest DCL and Spain is still the lowest. The difference across venues in average DCL as a fraction of trade size is now fairly small with, if anything DCL being larger on the primary exchange. Looking at pairwise average DCL levels, it is clear that DCL on alternative venues when a trade occurs on the primary tends to be much smaller than DCL 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 DCL measure. The DCL 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 DCL measure. However, some of these trades may have been executions of genuine duplicated 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 DCL measure represents a lower bound on true duplicated liquidity. To provide an upper bound, we also compute summary statistics for a DCL 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 duplicated limit orders. The adjustment roughly doubles the level of DCL measured as a fraction of quantity from just over 4% (measured across all stocks) to almost 9%. On some markets and some venues, DCL reaches 15%. Thus, factoring volume executed on the quote venue into the DCL definition significantly increases the scale of DCL. Performing the same adjustment to our DCL measure based on trade size leads to statistics in which DCL rises from roughly 20% to 25%. Thus, there is a rise here too, but proportionately less big.[11]

Table 5 about here

Returning to the original DCL measure, we proceed to investigate the variation of DCL with stocks' activity levels. Table 5 displays the average level of DCL per market value tercile. Differences in DCL expressed as a fraction of pre-trade liquidity are in general not very large, but there is a tendency for DCL to rise with market capitalization. This tendency is much clearer when DCL is expressed related to the size of the triggering trade. The table also demonstrates that DCL is negatively related to volatility in a stock, likely because when volatility is high, a market-making strategy which leaves multiple orders exposed on various venues is very costly. Finally, as one

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

would expect, the last panel of Table 4 shows us that DCL is larger on average for stocks with more fragmented trading. Presumably, when volume is dispersed across venues, liquidity suppliers find it profitable to provide consistent liquidity across those venues through order duplication.

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 DCL measure and those variables. They also find that the likelihood of duplicate orders being canceled also tends to rise with market cap and fragmentation. Thus, their results and ours are consistent.

Regarding the re-supply of liquidity in the 10ms following DCL, average refill rates on the same side are negative and average rates of liquidity provision on the opposite side of the market are positive for all stock terciles except two: the smallest stocks and the least fragmented, i.e., those with the lowest DCL.

Table 6 about here

It is important to understand whether DCL is mainly due to some categories of members. Table 6 decomposes average DCL 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 in two cases: (1) when comparing traders by speed and (2) when comparing members acting as principal with those acting for their clients.

When comparing traders by speed, as we would expect, the average DCL for HFTs is, at 5.75% of their total pre-trade liquidity, about 1.5 times larger than the average DCL associated with algo traders (AT) which is, in turn, around 1.4 times larger than DCL from slow traders. Thus, HFT trading strategies involve greater duplicated liquidity. ESMA (2016) report a similar finding for their direct analysis of order duplication. Further, strategies of liquidity re-supply immediately after canceling duplicated orders significantly differ by trading speed categories. First, fast traders re-supply between three and four times more liquidity than slow traders in the next 10ms. Second, while ATs re-supply liquidity on both sides of the market, HFTs re-supply on the opposite side only, with the highest rate of liquidity provision on the opposite side among all speed categories (3.08%, i.e., twice that of ATs and 4.6 times that of slow traders). An interpretation of these findings is that HFTs are actively engaged in cross-venue market-making. Their technological advantages permit them to link venues together through their liquidity supply decisions as they can manage the risk of having multiple versions of the same order live simultaneously on different venues.

When comparing members by trading capacity, DCL is typically higher when members are acting as principal rather than agent. This feature is strengthened by the fact that members acting as agent have both the greatest refill rate (3.16%) and the lowest liquidity provision rate on the opposite side (0.01%). Conversely, members acting as principal have low refill rates but they are more likely to add liquidity on the opposite side. Again, this supports the intuition that DCL activity is related to market-making strategies.

Let us recall that the starting point of a DCL 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 DCL is measured. For that reason, we also provide, in Table 6, the percentage of trades for which there is order duplication on the quote 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 DCL 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 DCL reaches one half of that of members trading as principal.

The fact that, on average, DCL differs systematically across member categories suggests that it may be important to control for such categories in our multivariate analysis. Also, the fact that member categories with the highest DCL levels - namely global traders, HFTs, and members acting as principal - re-supply more liquidity on the opposite side of the market that on the same side in the next 10ms, suggests that we should expect DCL to play a role in cross-venue market-making strategies. This will be considered in our analysis of DCL determinants. We now turn to our empirical model and identification strategy.

6. Determinants of duplicated liquidity

In this section we set out to identify the determinants of DCL. 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, optimal market-making strategies will likely entail posting duplicated orders on multiple venues. In this scenario, cancellations associated with duplicated orders are frictions associated with this market making activity.

To develop this idea, a basic high-speed market making strategy in an order book requires, first, the placing of limit orders on one side of the book (phase 1), and second, after execution of one of those orders, unwinding the position by placing limit orders on the other side (phase 2). Thus, one earns the bid-ask spread. It is valuable if the whole strategy is completed within a very short time interval so as to minimize risk and to enable repetition of the strategy as many times as possible within a day. Execution speed is thus key to maximizing the profits expected from such strategies. In fragmented markets, a trader operating a high-speed market-making strategy cannot know with certainty on which venue liquidity traders will appear first. Therefore, duplicating limit orders across books, with the intention to cancel residual orders as soon as the desired quantity (i.e., DCL) is executed in one book, increases expected market-making profits by reducing both execution delays and non-execution risk.

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. Thus we expect DCL to increase with fragmentation (Hypothesis H1) in line with more active market making by global members.

The incentive to post duplicated orders and the resulting DCL is greater when other options to improve execution probability, such as competing on price, are not available. DCL 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) and market making incentives to capture rents are greater (an argument similar to that in Yueshen, 2021). For that reason, we expect DCL to increase in the tick size (Hypothesis H2).

As duplicated liquidity is a tool used to increase the profits of limit order traders in their market making activities, we expect frequent liquidity suppliers (Hypothesis H3) and traders acting as principal (Hypothesis H4) to post more DCL than otherwise similar traders.

The eagerness of a liquidity supplier to trade further depends on her pre-trade inventory level. An inventory that deviates significantly from its optimal level gives a greater incentive to seek execution speed by posting DCL. Thus, we hypothesize that the DCL posted by a market member increases with the deviation between her stock inventory and its normal level (Hypothesis H5). H5 can also be interpreted as implying that we expect DCL to be greater when a liquidity supplier reverses her trading position (phase 2 in the market-making strategy) than when she builds it (phase 1 in the market-making strategy).

Fast traders are more likely to be actively making markets than slow traders through duplicated orders. To handle the possible friction of over-trading (i.e., the risk of being executed at multiple locations such that total quantity traded exceeds desired quantity) inherently linked to market making strategies, we expect the DCL of a market member to increase with her trading speed advantage (Hypothesis H6). That is, we expect fast traders to exhibit greater DCL than slow traders. 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 SORs to hit her limit orders on several platforms simultaneously. We thus posit that DCL 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 DCL to be greater on alternative platforms (Hypothesis H8).

We test these eight hypotheses by conducting a panel regression analysis of data measuring the DCL 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 DCL is not the result of shifts in liquidity by the same member from the quote venue towards the trade venue. We do so by computing the added liquidity on the trade venue as well as a DCL consolidated across platforms.

6.1 Global members

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

Our regression model is

$$
\begin{aligned}
&DCL_{tv\to qv}\left(\Delta t;i;t;m\right) \\
&=\alpha+\beta_1 FRAG_{i,t-1}+\beta_2 TICK_{i,t}+\beta_3 AGENT_{i,m}+\beta_4 LS_{i,m}+\beta_5 INV_{i,t-1,m} \\
&\quad+\beta_6 HFT_{i,m}+\beta_7 AT_{i,m}+\beta_8 SOR_{i,t-1}+\beta_9 SOR_{i,t-1}^2+\beta_{10} PEtoALT_{tv,qv}+\beta_{11} ALTtoPE_{tv,qv} \\
&\quad+\gamma_1 DCL_{HFT\backslash i,t,m}^{Others}+\gamma_2 DCL_{AT\backslash i,t,m}^{Others}+\gamma_3 DCL_{Slow\backslash i,t,m}^{Others} \\
&\quad+\delta_1 VOLUME_{i,t}+\delta_2\sigma_{i,t-1}+\delta_3 PRICE_{i,t}+\delta_4 TRADESIZE_{i,t,m} \\
&\quad+\mu_1 IMB_{i,t}+\mu_2 IMB_{i,t-1}+\varepsilon_{i,t,m,tv,qv}.
\end{aligned}
\tag{5}
$$

$DCL_{tv\to qv}\left(\Delta t;i;t;m\right)$ is the aggregated DCL on venue *qv* resulting from a trade on venue *tv*, for stock *i*, in 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 DCL 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.[12] $TICK_{i,t}$ 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. 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 absolute value. The measure thus represents the distance between current inventory and its 'normal' level for that member and stock.

[12] 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.

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 DCL.

$SOR_{i,t}$ 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 engaged 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 DCL (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;

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

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 duplicated 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 deliver 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 DCL (i.e., quote venue *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 DCL by other HFT members ($DCL^{Others}_{HFT\backslash i,t,m}$), other AT members ($DCL^{Others}_{AT\backslash i,t,m}$), and other slow traders ($DCL^{Others}_{Slow\backslash i,t,m}$) excluding member m (denoted by \\$m$) in period t for stock i. We also include a set of stock-time characteristics that are determinants of liquidity. 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 DCL 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 DCL 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 excluded when the left-hand side variable is DCL 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 DCL 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 DCL 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 duplicated liquidity (DCL=0), or all duplicated liquidity is withdrawn (DCL=1). The last column in Table 7 presents the results where we scale DCL by the trade size at the trade 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 DCL than otherwise similar trades against slow traders (the base case), and that HFTs post more DCL than ATs, with a statistical significance at the 1% level. Based on the first column of the table ($\Delta t = 10$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 DCL than AT members. DCL 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. However, the economic magnitude implied by the estimated coefficients is relatively modest, with a one standard deviation increase in inventory leading to a fall in DCL of around 0.1 percentage points. So more extreme inventory positions are associated with a slightly smaller DCL. This suggests that we should reject our hypothesis H5. Members do not use DCL to manage inventory in times when inventory is extreme. Instead, the sign of the coefficient is consistent with members being more comfortable using DCL strategies when their inventories are relatively small. In other words, limit order traders are more likely to use DCL in the first phase of their market making strategy, i.e., when they are happy to build trading positions, rather than in the second phase, when they are actively trying to unwind large positions. In that second phase, probably because of over-execution risk, it is likely that DCL is more costly.

The last column in Table 7 presents the results where we scale DCL by the trade size at the trade venue *tv*. It allows us to assess what fraction of the trade size executed on the trade venue *tv* is withdrawn by members on the quote 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. Again, this is consistent with H6, and one interpretation of this is that fast traders are more active in making markets than slow traders. 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 those 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 DCL. Members have more incentives to cancel orders when trade size on the trade venue is larger. Results for Δt = 10ms (first column) show that when trade size doubles, DCL increases by 1.2 percentage points.

The next rows in Table 7 show the results for the "platform characteristics". Based on column 1 ($\Delta\tau$ = 10ms), the *PEtoALT* coefficient shows that DCL is 1.8 percentage points less pronounced when the trade takes place on the primary exchange and the quote venue (where DCL is measured) 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, DCL 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 canceled on the primary exchange. This is in line with H8 and with cross-platform market making.

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

Across the various time windows, the significant positive coefficients on trading volume and fragmentation imply that DCL 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 DCL 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 DCL, which is not in line with an information-based interpretation. DCL 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 DCL might be more intensively used when undercutting by price is more difficult.

Finally, there is a concave relationship between smart order routing and DCL. This generates small increases in DCL 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, DCL would be 4 percentage points greater than if SOR was zero, while if SOR was at 80% of trading, DCL would be almost 25 percentage points lower). So, when smart order routers are used extensively, we see low use of DCL, likely due to the multiple execution risk that SOR technology exposes the users of DCL to (in line with H7.)

6.2 Member categories

Table 8 shows the results of Equation (5), where $\Delta t = 10$ms, for subsamples that focus on various member categories. This allows us to study whether specific 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 quote venue than ATs (i.e., the base case) following a trade on the trade venue. Compared with the first column of Table 7, which presents results for all member categories, some interesting differences in the coefficient signs and magnitudes 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. The negative coefficients of both *PEtoALT* and *ALTtoPE* in the case of HFTs (Column (3) of Table 8) indicate that the level of DCL posted by HFTs is the highest when both exchanges involved in their duplication strategies are alternative. Second, co-movement of DCL is most pronounced among own-member types. Columns (2) and (3), for example, show coefficients of around 0.15 for DCL_{AT}^{Others} and 0.20 for DCL_{HFT}^{Others}, respectively. These are considerably larger than the coefficients on other member categories.

Regarding order flow characteristics, trader inventory, and stock volatility, the coefficients of Table 8 are consistent in sign with those of Table 7. The coefficients on tick size are significantly negative in only two of the five regressions of Table 8, namely those corresponding to liquidity suppliers. he negative relation found between DCL and tick size is statistically insignificant for other traders.

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

6.3 Alternative explanations: Is Duplicated-then-Canceled Liquidity really canceled?

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

To study this alternative explanation, we first check whether orders canceled on the quote venue (DCL) 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 DCL measured on the quote venue is also canceled 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) evolves in the time window before (i.e, at t) to after (i.e., to t+10ms) the trade taking place on a particular trade 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 we control for trades against our member in the event window. We find that, on average, the liquidity withdrawn from all books immediately after a trade equals 6.62% of a member's consolidated depth and 59.09% of trade size. Since these numbers are larger than our single-venue DCL measures (4.04% and 19.67%, respectively), we find that members are not shifting limit orders to the trade venue. Instead, further liquidity is being withdrawn from the trade venue.

Finally, for this section, we build on subsection 5.2 and study whether orders that are canceled in the consolidated order book are refilled within the 10ms following the time window over which DCL is measured (i.e., we analyze the "refill rate"). On average, we find a negative refill rate of -2.84% of the globally canceled liquidity of that member, indicating that members continued canceling liquidity in the 10ms after the trade interval.

7. Impact of the cancellation of duplicated limit orders on trading costs

Finally, we analyze how strategies involving the cancellation of duplicated orders affect the trading costs of liquidity takers (LTs). More specifically, we investigate the impact of DCL on effective spreads for various LT groups. We choose to focus our analysis on the DCL of HFTs, since our preceding results imply that HFTs are the drivers of DCL. On the one hand, we might expect markets where more duplicated orders are submitted to be those in which 'genuine' liquidity is harder to measure and so execution cost management might be more difficult. This may result in additional costs of trading. But on the other hand, given that DCL has been shown to contribute to more efficient cross-venue market making, it may help to reduce trading costs.

We test these hypotheses by running panel regressions of the effective spreads paid by LTs on the level of DCL submitted by HFTs in the previous period. The analysis is conducted on measures of spreads and DCL aggregated by 15-minute period, stock, and venue. For each 15-minute period and stock, we compute the average DCL of HFTs, 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:

$$\begin{aligned} ES_{i,t,k} = a + b\,\sigma_{i,t} + cVOLUME_{i,t,k} + d\ PRICE_{i,t} + eTRADESIZE_{i,t,k} + fPE_{i,k} \\ + gD\hat{C}L^{HFT}_{i,t-1,k} + hES_{i,t-1,k} + \varepsilon_{i,t,k}. \end{aligned} \tag{6}$$

In this equation, the dependent variable $ES_{i,t,k}$ is the average effective spread for stock i in 15-minute period t on venue k. Our variable of interest is $D\hat{C}L^{HFT}_{i,t-1,k}$, which represents the amount of duplicated orders that are voluntarily canceled by HFTs for stock i, on venue k, over the previous 15-minute period. We take the DCL of the previous period with the intention of identifying causal effects of DCL on spreads. We may however still have an endogeneity problem as the decision to duplicate limit orders across venues in period $t-1$ might be determined by liquidity factors at play at both $t-1$ and t. To address this concern, we model the DCL of HFTs for stock i, on venue k, at period t in a first-stage regression specified as follows:

$$\begin{aligned} DCL^{HFT}_{i,t,k} = \alpha_0 + \alpha_1 DCL^{HFT}_{-i,t,k} + \alpha_2 DCL^{HFT}_{i,t-1,k} \\ + \beta_1 FRAG_{i,t-1} + \beta_2 TICK_{i,t} + \beta_3 SOR_{i,t-1} + \beta_4 SOR^2_{i,t-1} \\ + \gamma_1 VOLUME_{i,t} + \gamma_2 \sigma_{i,t-1} + \gamma_3 PRICE_{i,t} + \eta_{i,t,k}. \end{aligned} \tag{7}$$

Following Hasbrouck and Saar (2013) and Degryse *et al.* (2015), we use two instrumental variables in Equation (7): (1) the DCL of HFTs of all stocks in the same size tercile, stock i excluded , on the same venue, over the same period, denoted $DCL^{HFT}_{-i,t,k}$, and (2) the lagged DCL of HFTs for stock i on the same venue. Other variables are factors that have been demonstrated to determine DCL in Section 6, namely order flow fragmentation, tick size, SOR activity, total trading volume, price range, and price level, respectively measured by variables *FRAG*, *SOR*, *VOLUME*, σ, and *PRICE* in the same way as in equation (4). We include stock fixed effects and intraday time fixed effects identifying each 15-minute period of the trading session. The values of $DCL^{HFT}_{i,t,k}$ predicted by equation (7), denoted $D\hat{C}L^{HFT}_{i,t,k}$, are then used with a lag as the main regressor in equation (6).

In equation (6), we control for widely acknowledged determinants of spreads, such as: volatility, $\sigma_{i,t}$, measured as in equation (4) as the 15 minute price range; log trading volume $VOLUME_{i,t,k}$, computed as the log of euro trading volume for stock i on venue k over period t;

and the same price level measure, $PRICE_{i,t}$, that we used in equation (4). We also control for trade size and exchange type by introducing $TRADESIZE_{i,t,k}$, the average size of the trades that were used to construct the effective spread variable, and $PE_{i,k}$, a dummy equal to one when the venue for which we are computing effective spreads is the primary exchange, zero otherwise.

Furthermore, in order to examine whether the impact of the DCL of HFTs on effective spreads differs between the primary exchange and alternative venues, we modify equation (6) by interacting $D\hat{C}L^{HFT}_{i,t-1,k}$ with exchange dummies in the following way:

$$\begin{aligned} ES_{i,t,k} = a + b\,\sigma_{i,t} + cVOLUME_{i,t,k} + d\ PRICE_{i,t} + eTRADESIZE_{i,t,k} + fPE_{i,k} \\ + g_1 D\hat{C}L^{HFT}_{i,t-1,k} \times PE_{i,k} + g_2 D\hat{C}L^{HFT}_{i,t-1,k} \times ALT_{i,k} + hES_{i,t-1,k} + \varepsilon_{i,t,k}, \end{aligned} \tag{8}$$

where dummy $ALT_{i,k}$ equals one when venue k, where effective spreads are measured, is an alternative exchange, zero otherwise. In other words, $ALT_{i,k}$ simply equals $1-PE_{i,k}$. g_1, the coefficient of the product of the DCL measure and the primary exchange dummy, identifies the impact of the DCL of HFTs on the effective spreads of the primary exchange, while g_2, the coefficient of the product of *DCL* and *ALT*, identifies the impact on the spreads of alternative exchanges.

In both equations (6) and (8), autocorrelation is accounted for by including the first lag of the dependent variable, and intraday time fixed effects identifying each 15-minute period of the trading session are included. We run three versions of those two regressions by computing average effective spreads for three groups of LTs, differentiated by trading speed. Those three groups are slow liquidity takers, algorithmic liquidity takers, and HFT liquidity takers, respectively.

Table 9 contains estimates of our second-stage regressions (6) and (8). There are several familiar results in the table (e.g., spreads decrease with volume and price level, increase with volatility and trade size, and are positively autocorrelated). We also see that they are, on average, larger on the primary exchange.

Table 9 about here

As for the coefficients on DCL, we find that, when looking at all venues combined, greater DCL of HFTs leads to larger effective spreads for slow LTs and smaller effective spreads for fast LTs at a 1% level of statistical significance. The adverse effect on the trading costs of slow LTs is most severe on the alternative venues and is smaller on the primary exchange, where its economic significance is greatly reduced, and its statistical significance is at the 10% level only. Presumably, the DCL of HFTs is less detrimental to the effective spreads of slow traders on the primary exchange because slow traders themselves form a greater part of the trader population there, and thus the DCL of a subgroup of traders has smaller impact. Further, slow traders are typically less active on alternative exchanges. We should therefore consider the g_1 coefficient as more relevant than the g_2 coefficient for this group of traders. In this regard, the adverse effect of HFTs' DCL on the trading costs of slow LTs can be seen as small in both economic size and statistical significance.

Regarding the magnitude of the beneficial effect for fast traders, the economic impact, all venues considered, is similar for ATs and HFTs, with the g coefficient having similar values for both trader categories. However, when differentiating between the primary exchange and alternative venues, the picture changes. Whereas HFTs benefit from DCL everywhere, be it on the primary exchange or on alternative venues, ATs benefit from it on alternative exchanges only. On the contrary, their trading costs increase with the DCL of HFTs on the primary exchange. Thus, taken together, ATs may benefit or not depending upon their activities across primary and alternative exchanges. Weighing up algo LTs' costs and benefits, the economic significance of the benefits on alternative exchanges is still larger than that of the cost increase on primary exchanges.

All in all, these findings, in conjunction with those of Sections 5 and 6, suggest that DCL makes market-making strategies more effective on the exchanges where HFTs are the most active relative to other traders, i.e., the alternative venues. Those more effective market-making strategies benefit all fast traders on alternative exchanges while they only benefit HFT liquidity takers on primary exchanges. The DCL of HFTs on primary exchanges adversely impact the trading costs of both slow and algo LTs, with ATs being more significantly affected than slow LTs. Considering that slow traders mainly trade on primary exchanges and that the significance of the adverse effect of DCL on their trading costs is weak there, we can conclude that the adverse effect of the cancellation of duplicated orders by HFTs is somewhat limited for slow LTs. In contrast, the fact that the cancellation of duplicated limit orders by HFTs on primary exchanges is more harmful for the trading costs of algo LTs than for those of slow traders comes, to some extent, at a surprise. It opens questions for future research on the differences between algo and HFT trading strategies and their interactions on primary exchanges.

8. Conclusion

The objective of this paper is to assess the scale of Duplicated-then-Canceled Liquidity (DCL) and to examine its determinants. Our DCL measure 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 limit orders on other venues. Such liquidity provision strategies are natural and valuable in a fragmented trading landscape where some liquidity suppliers are market-making across venues. Thus, on the one hand, duplicating orders across venues may benefit cross-market liquidity by improving execution probabilities and by bringing liquidity to the single-market players of all venues, but on the other hand, DCL may lead some market participants to over-estimate the true liquidity available in the marketplace.

By drawing on a unique dataset that covers 91 European stocks listed in nine different countries, and their trading on their respective primary exchanges plus the three main alternative trading venues in Europe, i.e., Chi-X, BATS, and Turquoise, we find that DCL is an economically significant phenomenon that deserves attention from market participants and regulators. Limit order duplication is however not always DCL. 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 DCL, this average percentage being greater on alternative venues (between 6% and 7%) than on primary exchanges (3.43%). Those figures are unlikely to be large enough either to counterbalance the depth improvement related to fragmentation reported in previous literature (e.g., Degryse et al. (2015), Foucault and Menkveld (2008), and Gresse (2017)), or to create significant 'noise' in total liquidity measures. Furthermore, DCL does not necessarily affect all traders in the same way, as fast traders using properly calibrated smart order routers may catch DCL before it is withdrawn.

DCL may however reach substantial levels for some stocks, platforms, or traders. The cross-section of our sample shows that DCL is greater for larger, more fragmented stocks, and for less volatile stocks. Further, DCL increases with trading volumes and trade size. It decreases when smart order routing is particularly prevalent. HFTs, traders acting as principal, and traders implementing multi-market market-making strategies post more DCL than others. Further, regarding HFTs, their use of DCL is the highest when they duplicate limit orders across alternative platforms. Those results are robust to changes in the time window used to measure DCL, and they are not significantly impacted by cancellations due to quote updating in response to trades. These results together suggest that DCL is used by traders implementing market-making strategies to serve disconnected liquidity pools in a consistent way (e.g., when fragmentation is large), that HFTs contribute substantially to this activity, and that, when market conditions do not favor duplication (e.g. when volatility is high), the use of the strategy declines. Further, when the liquidity demand side of the market is using tools that allow them to access all venues simultaneously, e.g. smart order routing, DCL reduces as it becomes risky, and potentially costly, for liquidity suppliers.

In our final piece of analysis, we find that while DCL leads slow traders to suffer marginally increased trading costs, DCL significantly reduces the execution costs of fast liquidity takers (i.e. ATs and HFTs) on alternative trading venues. We interpret the latter as evidence that DCL contributes to the efficiency of market-making strategies that HFTs and others run on alternative exchanges, allowing counterparties access to better prices on average.

Overall, we show that duplicated-then-canceled liquidity is a significant phenomenon in European equity markets. A downside of its use is that naive measurement of consolidated liquidity may overestimate true liquidity in fragmented electronic markets. However, our final results suggest that it is an important tool in facilitating cross-venue market-making, leading to improved realized liquidity on alternative venues and greater true consolidated 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.

Harris, Lawrence (1993). Consolidation, Fragmentation, Segmentation and Regulation. *Financial Markets, Institutions & Instruments*, 2(5), 1-28.

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. (2021). Queuing uncertainty of limit orders. 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. Market shares by 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 DCL as a percentage of quantities available in the book

		10ms	Liquidity re-supply in the next 10ms: Same side (Refill rate)	Liquidity re-supply in the next 10ms: Opposite side	20ms	50ms	100ms
Panel A - By country							
Belgium		5.81%	0.35%	1.17%	5.85%	5.93%	6.13%
France		4.95%	0.25%	0.72%	5.11%	5.31%	5.39%
Germany		3.04%	-0.35%	0.93%	3.44%	3.68%	3.73%
Ireland		3.26%	-0.74%	0.20%	2.59%	2.09%	2.32%
Italy		1.73%	-1.34%	1.05%	2.05%	2.19%	2.32%
The Netherlands		6.18%	-0.39%	0.92%	6.34%	6.20%	6.36%
Portugal		3.53%	0.12%	0.90%	3.60%	3.55%	3.51%
Spain		0.23%	0.06%	0.20%	0.43%	0.61%	0.50%
The United Kingdom		6.83%	-0.74%	0.55%	6.92%	6.81%	6.91%
All stocks		4.04%	-0.34%	0.73%	4.20%	4.26%	4.34%
Panel B - By platform							
Primary exchange		3.43%	-0.55%	0.61%	3.54%	3.48%	3.54%
Chi-X		6.58%	-0.78%	1.58%	7.06%	7.47%	7.61%
Turquoise		5.98%	0.58%	0.82%	6.22%	6.54%	6.72%
BATS		6.92%	-0.54%	0.56%	7.56%	8.19%	8.51%
Panel C - By pair of platforms							
Quote venue	*Trade venue*						
Primary exchange	Chi-X	3.74%	-0.48%	0.69%	3.87%	3.92%	4.02%
	BATS	1.96%	-0.19%	0.37%	2.00%	1.69%	1.50%
	Turquoise	3.30%	-0.57%	0.49%	3.38%	3.34%	3.37%
Chi-X	Primary exchange	6.61%	-0.86%	1.31%	7.11%	7.58%	7.80%
	BATS	5.25%	-1.03%	0.98%	5.56%	5.48%	4.97%
	Turquoise	6.31%	-0.31%	1.90%	6.51%	6.63%	6.60%
BATS	Primary exchange	6.19%	-0.68%	0.46%	6.82%	7.54%	7.93%
	Chi-X	8.50%	-1.41%	1.00%	9.39%	9.77%	9.72%
	Turquoise	8.55%	-0.86%	0.66%	8.79%	9.02%	9.21%
Turquoise	Primary exchange	5.86%	0.65%	0.67%	6.07%	6.45%	6.73%
	Chi-X	5.99%	-0.33%	1.27%	6.28%	6.34%	6.30%
	BATS	4.94%	-0.89%	0.99%	5.13%	5.03%	5.17%

This table reports statistics on DCL measured as a percentage of quantities available in the order book of the quote venue prior to executions on the trade venue. Means of DCL 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 DCL at the 10ms horizon, the table also reports average rates of liquidity re-supply within the next 10ms (refill rates on the same side of the market and liquidity provision on the opposite side). Those rates are expressed as a percentage of DCL and winsorized at the 99% level. DCL and liquidity re-supply 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 DCL 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					
Quote 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 DCL measured as a percentage of the size of the trade that triggers the DCL measurement. Means of DCL are presented by country (Panel A), platform (Panel B), and pair of platforms (Panel C), for different time windows (10ms, 20ms, 50ms, and 100ms). DCL 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 DCL values are averaged for each stock over the entire month and equally-weighted means across stocks are calculated.

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

		Average DCL as a % of pre-trade liquidity (10ms)	Liquidity re-supply in the next 10ms		Average DCL as a % of trade size (10ms)
			Same side (Refill rate)	Opposite side	
Market value tercile	**Market value range (EUR Mn)**				
1	195 to 1,833	3.45%	0.39%	0.27%	16.17%
2	1,989 to 5,846	3.86%	-0.50%	0.58%	17.98%
3	6,152 to 118,942	4.79%	-0.88%	1.32%	22.42%
Volatility tercile	**Daily volatility range**				
1	0.0706% to 0.1253%	4.96%	-0.52%	0.51%	22.74%
2	0.1266% to 0.1549%	3.97%	-0.22%	0.57%	18.78%
3	0.1549% to 0.3266%	3.17%	-0.26%	1.10%	15.04%
Fragmentation tercile	**Fragmentation index range**				
1	1.0604 to 1.5520	1.68%	0.42%	0.29%	7.18%
2	1.5553 to 2.0663	3.35%	-0.27%	0.75%	15.10%
3	2.0831 to 3.0714	7.00%	-1.13%	1.14%	33.90%

This table reports statistics on DCL by market value tercile, volatility tercile, and fragmentation tercile. DCL is here measured both as a percentage of the pre-trade quantity posted by a member in the order book and as a percentage of the size of the trade that triggers the measurement of DCL. The table also reports average rates of liquidity re-supply within the next 10ms (refill rates on the same side of the market and liquidity provision on the opposite side). Those rates are expressed as a percentage of DCL and winsorized at the 99% level. DCL and liquidity re-supply 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 DCL as a percentage of pre-trade liquidity by member category

		Average DCL as a % of pre-trade liquidity (10ms)	% of cases with duplication	Liquidity re-supply in the next 10ms		Average DCL as a % of trade size (10ms)
				Same side (Refill rate)	Opposite side	
Trading aggressiveness	Liquidity Taker	3.69%	34.42%	1.34%	0.90%	13.53%
	Liquidity Supplier	3.81%	54.84%	-0.02%	2.29%	18.43%
Trading scope	Local	2.11%	3.31%	0.26%	0.59%	11.59%
	Global	3.80%	57.81%	0.38%	1.87%	16.50%
Trading speed	Slow	2.70%	32.60%	0.08%	0.67%	12.32%
	AT	3.76%	56.84%	0.81%	1.47%	12.52%
	HFT	5.75%	53.65%	0.09%	3.08%	16.87%
Capacity	Agent	1.94%	16.78%	3.16%	0.01%	5.48%
	Principal	3.93%	51.23%	0.42%	1.81%	17.56%

This table reports statistics on DCL, and liquidity re-supply rates (i.e, refill rates, and rates of liquidity provision on the opposite side) by member category. It also provides the proportion of trades for which pre-trade liquidity is duplicated. Liquidity re-supply rates are expressed as a percentage of DCL and winsorized at the 99% level. DCL and liquidity re-supply 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 DCL for global market members

DCL measure	DCL as fraction of pre-trade liquidity				DCL as fraction of trade size
Time window for DCL	10ms	20ms	50ms	100ms	10ms
Member characteristics					
HFT	0.0788***	0.0788***	0.0807***	0.0826***	0.2197***
	(0.0014)	(0.0015)	(0.0015)	(0.0015)	(0.0040)
AT	0.0280***	0.0279***	0.0284***	0.0287***	0.0547***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0039)
Agent	-0.0203***	-0.0198***	-0.0210***	-0.0231***	-0.0522***
	(0.0025)	(0.0025)	(0.0026)	(0.0026)	(0.0072)
Liquidity supplier	0.0258***	0.0275***	0.0280***	0.0281***	0.0851***
	(0.0012)	(0.0012)	(0.0012)	(0.0013)	(0.0034)
Average inventory $_{t-1}$	-0.0009***	-0.0008***	-0.0009***	-0.0010***	-0.0041***
	(0.0010)	(0.0010)	(0.0010)	(0.0011)	(0.0029)
Trade characteristics					
Trade size $_t$	0.0119***	0.0119***	0.0120***	0.0119***	
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	
Platform characteristics					
PE-to-alternative	-0.0183***	-0.0170***	-0.0149***	-0.0147***	-0.0611***
	(0.0014)	(0.0014)	(0.0014)	(0.0015)	(0.0040)
Alternative-to-PE	0.0267***	0.0269***	0.0275***	0.0284***	0.1100***
	(0.0014)	(0.0014)	(0.0014)	(0.0015)	(0.0039)
Other market member DCL					
DCL(other, HFT) $_t$	0.0472***	0.0480***	0.0497***	0.0495***	5.28E-06***
	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0000)
DCL(other, AT) $_t$	0.0540***	0.0458***	0.0513***	0.0392***	3.48E-05***
	(0.0021)	(0.0003)	(0.0019)	(0.0004)	(0.0000)
DCL(other, Slow) $_t$	0.0000	0.0000	0.0000*	0.0000*	1.39E-05***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Order flow characteristics					
Volume $_t$	0.0047***	0.0050***	0.0051***	0.0051***	0.0165***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0023)
Order imbalance $_t$	-0.0092***	-0.0098***	-0.0098***	-0.0106***	-0.0383***
	(0.0034)	(0.0034)	(0.0035)	(0.0035)	(0.0100)
Order imbalance $_{t-1}$	-0.0009	-0.0009	-0.0004	-0.0005	0.0009
	(0.0032)	(0.0032)	(0.0033)	(0.0033)	(0.0094)

Fragmentation $_{t-1}$	0.0020***	0.0023***	0.0024***	0.0023***	0.0060***
	(0.0013)	(0.0013)	(0.0013)	(0.0014)	(0.0037)
SOR $_{t-1}$	0.3755***	0.3873***	0.3989***	0.4122***	1.2192***
	(0.0275)	(0.0281)	(0.0289)	(0.0294)	(0.0850)
(SOR $_{t-1}$)^2	-0.8473***	-0.8871***	-0.9174***	-0.9625***	-3.1961***
	(0.1388)	(0.1433)	(0.1474)	(0.1506)	(0.4605)
Stock characteristics					
Price range $_{t-1}$	0.0266***	0.0233***	0.0223***	0.0238**	0.0473*
	(0.0499)	(0.0515)	(0.0533)	(0.0544)	(0.1378)
Price	-0.0053	-0.0022	-0.0051	-0.0027	0.0092
	(0.0201)	(0.0203)	(0.0208)	(0.0211)	(0.0579)
Tick	-15.1442***	-14.2076***	-12.3683**	-8.1827	-128.6595***
	(27.8036)	(28.2094)	(28.8822)	(29.2088)	(86.0085)
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 DCL by member, stock, and pairs of platforms on various factors. DCL 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. DCL 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 quote venue, i.e., the venue where the member's liquidity is potentially withdrawn. Reported coefficients are the marginal effects of the explanatory variables on DCL, conditional on DCL being positive. The independent variables of interest comprises 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 suppliers; the member's lagged average standardized absolute inventory, a PE-to-alternative dummy equal to one when the trade venue is the primary exchange and the quote venue an alternative platform; an alternative-to-PE dummy equal to one when the trade venue is an alternative platform and the quote venue is the primary exchange, a lagged fragmentation index, and the lagged and the contemporaneous values of a SOR proxy. The control variables include a measure of daily realized volatility; the lagged and the contemporaneous values of 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 DCL measured for other HFT members; the contemporaneous DCL measured for other AT members; and the contemporaneous DCL measured for other slow traders. ;. When DCL is measured as a fraction of pre-trade quantities in the book of the quote venue, the average size of the trades triggering the DCL observation is also controlled for, and the Tobit specifications are double-censored with a lower bound set to 0 and an upper bound set to 1. DCL 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 DCL 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***
	(0.0015)			(0.0017)	(0.0016)
AT				0.0332***	
				(0.0017)	
Agent	-0.0161***	-0.0075***			
	(0.0035)	(0.0046)			
Liquidity supplier	0.0280***	0.0440***	0.0199***		
	(0.0016)	(0.0028)	(0.0020)		
Average inventory $_{t-1}$	-0.0020***	-0.0016***	-0.0026***	-0.0011***	-0.0023***
	(0.0012)	(0.0021)	(0.0014)	(0.0011)	(0.0013)
Trade characteristics					
Trade size $_t$	0.0112***	0.0157***	0.0045***	0.0110***	0.0106***
	(0.0007)	(0.0012)	(0.0010)	(0.0007)	(0.0008)
Platform characteristics					
PE-to-alternative	-0.0240***	-0.0132***	-0.0363***	-0.0080***	-0.0060***
	(0.0018)	(0.0028)	(0.0022)	(0.0017)	(0.0020)
Alternative-to-PE	-0.0033***	0.0284***	-0.0430***	0.0137***	-0.0178***
	(0.0017)	(0.0029)	(0.0021)	(0.0016)	(0.0019)
Other market member DCL					
DCL(other, HFT) $_t$	0.0810***	0.0022***	0.1953***	0.0495***	0.0775***
	(0.0029)	(0.0044)	(0.0037)	(0.0026)	(0.0023)
DCL(other, AT) $_t$	0.0870***	0.1530***	0.0060***	0.0498***	0.0757***
	(0.0017)	(0.0026)	(0.0030)	(0.0016)	(0.0018)
DCL(other, Slow) $_t$	0.0000	0.0155***	0.0000	0.0000	0.0000
	(0.0000)	(0.0065)	(0.0000)	(0.0000)	(0.0000)
Order flow characteristics					
Volume $_t$	0.0058***	0.0075***	0.0047***	0.0030***	0.0033***
	(0.0010)	(0.0016)	(0.0012)	(0.0009)	(0.0010)
Imbalance $_t$	-0.0084***	-0.0060***	-0.0113***	-0.0089***	-0.0069***
	(0.0043)	(0.0068)	(0.0054)	(0.0038)	(0.0045)
Imbalance $_{t-1}$	-0.0036***	-0.0040***	-0.0044***	-0.0013*	-0.0022***
	(0.0040)	(0.0064)	(0.0051)	(0.0036)	(0.0042)
Fragmentation $_{t-1}$	0.0035***	0.0046***	0.0034***	0.0013***	0.0024***
	(0.0016)	(0.0026)	(0.0020)	(0.0015)	(0.0018)
SOR $_{t-1}$	0.3389***	0.4541***	0.1731***	0.3948***	0.3193***
	(0.0363)	(0.0561)	(0.0463)	(0.0316)	(0.0386)

$(SOR_{t-1})^2$	-0.9354***	-1.1113***	-0.5573***	-0.9429***	-0.8731***
	(0.1939)	(0.2846)	(0.2611)	(0.1595)	(0.2037)
Stock characteristics					
Price range $_{t-1}$	0.0377***	0.0490**	0.0356**	0.0300**	0.0240
	(0.0618)	(0.1074)	(0.0674)	(0.0602)	(0.0666)
Price	-0.0033	-0.0239***	0.0191**	0.0094*	0.0063
	(0.0242)	(0.0384)	(0.0299)	(0.0236)	(0.0264)
Tick	3.0240	-5.3349	5.8341	-31.6777***	-18.8915**
	(29.9785)	(52.9660)	(33.8387)	(33.3476)	(33.9912)
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 DCL by member, stock, and pairs of platforms on various factors. DCL is computed as a fraction of pre-trade liquidity on the quote venue over 10ms windows. DCL 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 quote venue, i.e., the venue where the member's liquidity is potentially withdrawn. Reported coefficients are the marginal effects of the explanatory variables on DCL, conditional on DCL being positive. The independent variables of interest comprises 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 suppliers; the member's lagged average standardized absolute inventory, a PE-to-alternative dummy equal to one when the trade venue is the primary exchange and the quote venue an alternative platform; an alternative-to-PE dummy equal to one when the trade venue is an alternative platform and the quote venue is the primary exchange, a lagged fragmentation index, and the lagged and the contemporaneous values of a SOR proxy. The control variables include the average size of the trades triggering the DCL observation, a measure of daily realized volatility; the lagged and the contemporaneous values of 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 DCL measured for other HFT members; the contemporaneous DCL measured for other AT members; and the contemporaneous DCL measured for other slow traders. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level respectively.

Table 9. Impact of the DCL of HFTs on the effective spreads of liquidity takers

	Effective spreads of slow LTs		Effective spreads of ATs/LTs		Effective spreads of HFTs/LTs	
Price range $_t$	6.94E-04***	6.92E-04***	4.94E-04***	5.00E-04***	6.43E-04***	6.42E-04***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Volume $_t$	-3.38E-05***	-3.40E-05***	-3.18E-05***	-3.17E-05***	-2.77E-05***	-2.77E-05***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Price	-6.80E-05***	-6.81E-05***	-7.24E-05***	-7.23E-05***	-5.59E-06***	-5.57E-06***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Trade size $_t$	1.08E-05***	1.07E-05***	7.12E-06***	7.47E-06***	-4.55E-06***	-4.60E-06***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
PE	6.33E-05***	7.73E-05***	2.15E-05***	-6.10E-06	5.01E-05***	6.09E-05***
	(0.0000)	(0.0000)	(0.0000)	(0.2530)	(0.0000)	(0.0000)
$\widehat{\text{DCL}}$ (HFT) $_{t-1}$	1.11E-04***		-8.20E-05***		-8.70E-05***	
	(0.0000)		(0.0000)		(0.0000)	
$\widehat{\text{DCL}}$ (HFT) $_{t-1}$ × PE		3.63E-05*		5.22E-05**		-1.41E-04***
		(0.0860)		(0.0410)		(0.0000)
$\widehat{\text{DCL}}$ (HFT) $_{t-1}$ × Alternative		1.37E-04***		-1.31E-04***		-6.99E-05***
		(0.0000)		(0.0000)		(0.0000)
Effective spread $_{t-1}$	0.6992***	0.6987***	0.5696***	0.5690***	0.3140***	0.3139***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Fixed effects						
15-min period fixed effects	YES	YES	YES	YES	YES	YES
Adjusted R^2	64.32%	64.23%	45.79%	45.82%	16.15%	16.17%

This table reports the 2nd-stage OLS regression results of effective spreads by stock-venue combination on the DCL of HFTs, by 15-minutes period, for alternatively slow liquidity takers, algo liquidity takers, and HFT liquidity takers. The DCL considered is the DCL of HFTs as predicted in a 1st-stage regression in which the instruments are the contemporaneous DCL of HFTs in other stocks of the same size tercile and the DCL of HFTs for the same stock in the previous 15-minutes period. Dummies "PE" and "Alternative" indicate whether the observation stems from respectively the primary exchange or an alternative venue. The control variables include price range, traded volume, 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.